Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data*

Stephen E. Fienberg and William M. Mason
University of Minnesota . University of Michigan

Technical Report No. 286

*This work was supported by Grant NIE-G-76-0094 from the National Institute of Education, U.S. Department of Health, Education and Welfare, and by a grant from the Horace H. Rackham Graduate School, University of Michigan. The opinions expressed do not necessarily reflect the positions or policies of the National Institute of Education. We gratefully acknowledge the outstanding research assistance of Michael Battaglia, the expert typing of Carol Crawford, and the comments of Robert D. Mare.

ABSTRACT

We formulate an age-period-cohort specification for discrete data. When ages and periods are evenly and equally spaced there is an identification problem. We explain the problem. Sampling models are indicated and the minimal sufficient statistics are given. The iterative proportional fitting and Newton-Raphson methods for maximum likelihood estimation are described for age-period-cohort specifications. Various formulations of discrete response variables are listed, and their special features explained. We also consider the identification problem when ages and periods are evenly spaced, with multiple age groups forming a span equal to that between successive periods. Degrees of freedom are provided for all models discussed. An extended analysis of an example illustrates the formal results of the paper.

INTRODUCTION

This chapter discusses and resolves several issues arising out of a particular form of cohort analysis. The exposition is in terms of discrete data analysis, both because it may help those working with discrete data and because many of the general issues involved in the form of cohort analysis which concerns us can be elucidated with special clarity in the discrete data context.

The version of cohort analysis we consider involves measurements on some dependent quantity conditional on the ages of the units of observation and the period at which the measurements were taken. We assume that replicated cross-sections are available, and, in general, that repeated measurements on the same units of observation across replications are not available. The dates of data collection define the periods, and for each cross-section, measurements on age, on the dependent quantity, and possibly on other variables, are available. Cohort membership is then defined by knowledge of age and the date or period at which that age was attained. The questions of interest for the analyst are whether age, period, and cohort simultaneously determine the endogenous quantity, and if so, how. We assume that whatever the explanation of the effect of cohort on the endogenous variable, no more direct measurement of the explanans than cohort membership itself is available for immediate purposes. We make similar assumptions about age and period. If more direct measures of the phenomena presumed to underlie the effects of age or period or cohort are available, then the problems which concern us do not occur.

For the methods we describe, we make no assumptions about the kind of phenomenon to be studied. The most important of the problems we discuss (identification) is formal, and has been encountered, explicitly or implicitly, in the analysis of the economic value of material objects (Hall, 1971), the study of the political proclivities of humans (Knoke and Hout, 1974), in epidemiology (Greenberg, et al., 1950), in developmental psychology (Baltes et al., 1976; Schaie, 1977), and in other areas. Indeed, the main problem we study can be found in research settings in which the notion of cohorts does not concern the investigators. In the social sciences, certain models for square occupational mobility tables (Goodman, 1972; Bishop, Fienberg and Holland, 1975, pp. 225-8, 320-4; Pullum, 1975) must resolve what is formally the same identification problem faced in cohort analysis, as Goodman (1975a) has observed.[1] Models for the analysis of status inconsistency (Blalock, 1966; Hope, 1971, 1975) are also essentially formally analogous to those for the age-period-cohort problem and again must overcome what amounts to the same under-identification. We suspect the problem recurs outside of the social sciences as well. Despite this pervasiveness of the formal issue, fruitful discussion of the complexities of cohort analysis depends somewhat on substantive considerations, for which purpose we shall focus on social phenomena. Hereafter age will refer to people's ages, period will refer to a time span during which data on the people were collected, and cohort will refer to a set of people born during a known time span.

Our point of departure is the article of Mason et al. (1973) which considers the identification problem for situations in which the dependent quantity is treated as a joint function of age, period and cohort membership. That article considers the class of specifications

$$Y = f(A_i, P_j, C_k) \tag{1}$$

in which A_i denotes the i-th of I ages, P_j denotes the j-th of J periods, and C_k denotes the k-th of K = (I + J - 1) cohorts. Expression (1) allows the effect of age, period and cohort membership on Y to be arbitrary. Mason et al. consider whether expression (1) is estimable (identified) and conclude that it is only if certain kinds of restrictions are made. In particular, they point out that the model

$$Y = \beta_0 + \beta_1 A + \beta_2 P + \beta_3 C + \varepsilon$$

in which observations are scored on Y, by their age for A, by the specific point of measurement for P, and their year of birth for C, is not estimable if A, P and C have been scaled such that A = P - C for all observations. Mason et al. then go on to consider the more general situation in which the effects of age, period and cohort can take any functional form up to the inherent complexity allowed by the fineness of detail in the collected data. They do this by considering expression (1) as a multiple classification model (Yates, 1934). For this model they show that a single equality restriction on a pair of effects suffices to identify the model. While the choice of the pair of effects to be equated does not affect the fit of the model to the data, it does affect the estimates of all of the parameters. Mason et al. show that an additional pair of equated effects results in a unique fit to the data. Models with one pair of equated effects are just-identified; models with more than one pair of equated effects are over-identified. Since different just-identified models fit the same data identically, a choice between them must be made on the basis of a priori reasoning. A choice between over-identified models, or between over-identified and just-identified models, can be made on the basis of their relative fit to the data and a priori reasoning.

Expression (1) is not general, because the effects of cohort membership are assumed constant over ages and periods. Equivalently, the effects of age are assumed constant over cohorts and periods, or the effects of period are assumed constant over ages and cohorts. That is, expression (1) assumes there are no age-period, age-cohort or cohort-period interactions affecting the endogenous quantity. This limitation of expression (1) has been discussed by Glenn (1976), Mason et al. (1976), and Knoke and Hout (1976). Glenn's view is that expression (1) is too restrictive for the useful analysis of most cohort problems as they are defined substantively in the social and behavioral sciences. For example, Glenn argues that it is often reasonable to suppose that the effects of cohort membership first increase and then decline as the individuals in a cohort age. Expression (1) does not allow for this or other more complex possibilities. The reason that it does not is that such complex effects are inestimable with the kind of data we have (replicated cross-sections). Glenn (1976) does not provide formal specifications for the more complex models he thinks are likely to be most appropriate. It may be that some of the effects he thinks are most reasonable to posit may be captured by an age-cohort model which excludes period, or that an age-period-cohort model reproduces the patterns he thinks it important to model, and does so in a fashion which permits substantively useful explanation. Mason et al. (1976) point out that for a given application expression (1) is a model, that models purposely simplify, and that they should not be discarded out of hand because they do so. There have been a number of seemingly helpful applications of expression (1) to substantive phenomena. The suitability of a model for a particular application is determined by a variety of considerations.

Methodological as well as substantive progress with cohort analysis depends in large part on model specification and on the determination of the estimability of the parametric structure of the specification. If, given the kind of data typically available for use in cohort analyses certain specifications are not estimable, then knowledge of this result is useful. Attempts to estimate these specifications are ostensibly precluded, and attention can be directed to the generation of new forms of data which do support estimation of the desired model, or to the creation of more direct *ex post facto* measures of the phenomena that age, period, or cohort membership are presumed to measure in the substantive contexts under consideration. Our goals here are different. We accept a specification which requires the form of the relationship between the dependent quantity and age, period, and cohort to be additive, and assume that the dependent quantity is a discrete variable. We then reconsider problems of identification and interpretation and develop the discussion more fully than was done by Mason et al. (1973). In the process, we clarify a number of points emerging from their discussion. For example, we show that the inestimability of the linear effects of age, period, and cohort effects noted by Mason et al. (1973) is precisely the reason that the age, period, and cohort effects in the multiple classification model they consider are also inestimable, and that despite the inestimability of the linear effects, the higher order effects are always estimable. We also provide new results, some of which are specific to the analysis of discrete data.

The remaining five sections are arranged as follows: (1) The Specification. This section deals with the basic specification and sampling models we wish to consider for discrete response variables. (2) The 3 × 3 × 2 Case. We first consider the special case of three age groups, three periods and a dichotomous response variable in order to develop the major points about identification, estimation and goodness-of-fit as clearly as possible before proceeding to more general cases. This section also illustrates for the 3 × 3 × 2 case how to arrange the data for computations with the two alternative algorithms (iterative proportional fitting and Newton-Raphson) discussed in the paper. (3) The I × J × 2 Case. This section generalizes the discussion of the 3 × 3 × 2 case to arbitrarily many age groups and periods. (4) The I × J × L Case. This section considers the case in which the response variable is polytomous. One of the interesting results of this section is that a polytomous response variable can sometimes be re-arranged into a series of asymptotically independent dichotomies, so that the data can be treated as a series of $I \times J \times 2$ problems. (5) <u>Differently Spaced Age and Period Intervals</u>. This section first considers the case in which the age groups and periods are evenly spaced but there are multiple age groups per period interval. A new identification problem crops up in this case. We indicate the solution to this problem, and generalize the earlier results for estimation and degrees of freedom to this case. We then consider the case of unevenly spaced periods; there is no currently satisfactory way to estimate age-period-cohort models in this instance. (6) <u>Example</u>. We develop an empirical example in some detail to illustrate the results obtained in preceding sections and the general nature of reasoning appropriate to the specification of age-period-cohort models. The example pertains to the educational attainment of white males, and uses data from the 1940, 1950, 1960 and 1970 U.S. Censuses.

THE SPECIFICATION

Parameter Structure: The Basic Logistic Response Model

We are interested in situations involving survey data for each of J evenly spaced points in time, and where for each such period we have our data broken down by the age group of the respondent, e.g., 20-24 years. We assume that the range in years covered by each age group equals the period interval, i.e., the interval in years between successive points in time for which we have data. Thus all those in a given age group, say i, at period j correspond to the same birth cohort as those in age group $i + 1$ at the subsequent period $j + 1$. If there are I age groups and J periods, then there are $I + J - 1$ cohorts. Note that within cohorts we do not necessarily follow the same individuals across time as in a panel study.

We are concerned with the simultaneous effects of age, period, and cohort on a categorical response variable. For simplicity we begin with a dichotomous response. Let $P_{ijk|1}$ denote the probability of a positive response given age i, period j, and cohort $k = i - j + J$, and let $P_{ijk|2} = 1 - P_{ijk|1}$ denote the corresponding probability of a negative response. We would like to represent some function of these response probabilities as being additive in the effects of interest. Because $P_{ijk|1}$ is a probability lying between 0 and 1, and because of the sampling schemes typically assumed to have been used to generate the data, a <u>natural</u>[2] as well as a convenient model to adopt is the linear logistic response model, which is based on the logarithm of the odds:

$$\Omega_{ijk} = \log\left(\frac{P_{ijk|1}}{P_{ijk|2}}\right) = \log\left(\frac{P_{ijk|1}}{1 - P_{ijk|1}}\right)$$

$$= W + W_{1(i)} + W_{2(j)} + W_{3(i-j+J)} \,, \qquad (2)$$

where the subscripted parameters in (2) are deviations[3] from W, i.e.,

$$\sum_i W_{1(i)} = \sum_j W_{2(j)} = \sum_k W_{3(k)} = 0 \quad . \qquad (3)$$

This model postulates simultaneous age, period, and cohort effects on the log-odds (or logit) of the probability of success. The notation we use here is consistent with that in Bishop, Fienberg, and Holland (1975) and Fienberg (1977). The model can be expanded to include further explanatory variables (e.g., sex, race) as well as their interactive effects with age, period, and cohort. The model is directly analogous to the age-period-cohort model for quantitative response variates examined by Mason et al. (1973).

The logistic response model of expressions (2) and (3) contains $2I + 2J - 3$ independent parameters, and we need to consider whether all of these parameters are estimable given the form of data. For model (2)-(3), as we point out below, some parameters cannot be estimated. In the cohort analysis literature this fact is known as the identification (or estimability) problem, and has been a stumbling block for those who wish to analyze archival data using age, period, and cohort membership in the same model.

Sampling Models

Two basic sampling models are appropriate for the estimation procedures we describe in this chapter. These are: (a) for each period we have survey data generated by a simple random sample of respondents; (b) for each period we have survey data generated by a stratified random sample. We could, of course, stratify by some other variable in addition to or in place of age, but then this variable would have to be formally included in our parametric structure and in the logistic response model of expressions (2)-(3).

Since most samples are drawn from finite populations and the formal sampling models are based on infinite populations, we make the usual caveat that the population size is sufficiently large that the difference between sampling with and without replacement is unimportant (i.e., we do not need to correct for the finite size of the population).

The sampling scheme appropriate for situation (a) is known as <u>multinomial</u> and that for situation (b) as <u>product-multinomial</u>. Bishop, Fienberg, and Holland (1975), Fienberg (1977), and Haberman (1974) all describe these two sampling schemes more formally and point out how they relate to each other mathematically, and for purposes of maximum likelihood estimation of parameters.

Few if any large scale surveys, such as those conducted by the U.S. Bureau of the Census or by non-Government survey organizations, use simple random samples. For example, the National Assessment of Educational Progress used a complex four-stage sample design involving both clustering and stratification (National Assessment of Educational Progress, 1972). Thus, the standard methods based on simple random sampling are not directly applicable, in part because they are insensitive to dependencies among sampling units. There have been several attempts to deal analytically with the analysis of discrete data from complex sample surveys (Altham, 1976; Cohen, 1976; Koch et al., 1973; Schuster and Downing, 1976), but these methods are at least as complex as the designs themselves and are difficult to implement except in special cases.

In complex sample surveys involving both clustering and stratification, standard errors for parameter estimates based on simple random sample assumptions typically underestimate the true standard errors. Thus one way to cope with complex sample designs without going into the analytical complexities is to "adjust" the overall sample size by a suitable fudge factor or "design effect," and then proceed as if simple random sampling had been used. Kish and Frankel (1974) discuss some aspects of this problem, and the use of design effects has been proposed for various large scale national surveys (Penick and Owens, 1976, p. 31). As a rough rule of thumb, practitioners often reduce the sample size by factors ranging from 10% to 50%.

In the example we consider in the final section of this chapter, the data for some of the cross-sections come from a complete census, and for the remainder the data come from 20% and 25% samples of the U.S. population. Clearly the assumption of simple random sampling makes no sense for the complete census, and the finite population correction cannot be ignored for the 20% and 25% samples. One way to resolve this difficulty is to think of the data as if they comprised a random sample from some hyperpopulation. In the context of our example, we are interested only in the U.S. population and the hyperpopulation ploy is unhelpful. In view of this difficulty, we use methods appropriate for simple random samples (with replacement) in the context of population and 20% and 25% sample surveys for descriptive purposes only, and thus, for example, we treat goodness-of-fit test statistics merely as indices of fit.

IDENTIFICATION, ESTIMATION, AND GOODNESS-OF-FIT: THE CASE OF 3 AGE GROUPS, 3 PERIODS AND A DICHOTOMOUS RESPONSE VARIABLE

The basic methodological problem is most simply illustrated in the context of data from 3 periods in time with information on 3 broad age groups, and for which the dependent variable is dichotomous. Corresponding to this structure are 5 birth cohorts.

The Basic Data Array

For the 3-age and 3-period case the basic data with which to assess the adequacy of model (2)-(3) would come from 3 surveys, one for each period. The data would consist of counts $\{x_{ijk\ell}\}$, where $\ell = 1$ corresponds to a positive response and $\ell = 2$ to a negative one, with the counts forming a $3 \times 3 \times 2$ cross-classification with the marginal configurations $\{x_{+j++}\}$ (these are the sample sizes for the independent simple random samples for each of the periods) fixed by design, as depicted in Table 1. The key thing to remember is that one of the first three subscripts is redundant since $k = i - j + 3$.

<u>Table 1 here</u>

Let $\{m_{ijk\ell}\}$ be the expected cell values corresponding to Table 1 under the logistic response model (2)-(3), with

$$m_{ijk1} = x_{ijk+}P_{ijk|1} \tag{4}$$

and

$$m_{ijk2} = x_{ijk+}P_{ijk|2} = x_{ijk+}(1 - P_{ijk|1}) \quad . \tag{5}$$

Then the basic logistic response model can be written in terms of expected cell values as

$$\log\left(\frac{m_{ijk1}}{m_{ijk2}}\right) = \Omega_{ijk} = W + W_{1(i)} + W_{2(j)} + W_{3(i - j + J)} , \qquad (6)$$

and analyses involving this model treat the marginal configuration $\{x_{ijk+}\}$ as fixed, even though only the totals $\{x_{+j++}\}$ are fixed by design.

Alternate Data Displays

If cohort effects in model (2)-(3) were absent, the data array of Table 1 could still be used, ignoring the third subscript. Removing cohort effects simplifies the logistic response model to one with no second order interaction involving the joint effect of age and period on the response. This model corresponds to a standard loglinear model and can be handled in a straightforward manner by the techniques described in Bishop, Fienberg, and Holland (1975) or Fienberg (1977).

One of the reasons we have trouble thinking about age, period, and cohort together is that we usually present data in the rectangular form of Table 1, i.e., as age by period. If in the analyst's view cohort effects are more important than age or period effects, then age by cohort or period by cohort tables may be preferable forms of display. Alternative forms of the positive response layer of the $3 \times 3 \times 2$ array in Table 1 are given in Table 2 (age by cohort) and in Table 3 (period by cohort). It is insufficient for the analyst to "eyeball" Tables 1-3 to decide the relative magnitudes of age, period, and cohort effects, but insights into their presence or absence can sometimes be gleaned from a simple inspection of these tables.

Tables 2-3 here

Tables 2 and 3 are incomplete contingency tables in which the dashes represent structural zeros--categories that are *a priori* impossible given the way in which the data are collected and restructured. For example, it is impossible to observe individuals from certain cohorts in surveys taken at times before the individuals are born.

The practice of examining archival data using all three forms of display might be illuminating, especially if only two of age, period and cohort are present. The major differences among the displays stem from the asymmetric fashion in which they reveal the three underlying variables. In Table 1 cohort effects represent the interaction between age and period; in Table 2 period effects represent the interaction between age and cohort; and in Table 3 age effects represent the interaction between period and cohort.

When we measure the cohort effects in terms of interaction contrasts (for a definition of interaction contrasts, see Graybill (1976, p. 564)) we are forced to confront the fact that we cannot separate the cohort dimension from a specific form of interaction between age and period. This is a conceptual problem that cannot be dismissed by mathematical fiat or by statistical slight-of-hand. Moreover, by thinking of cohort effects as interaction effects we take the main effects associated with the rows and columns of an age by period table as marginal to the interaction effects, i.e., the interaction effects are the residuals once we take out the main effects.[4] It turns out that when interaction is present in a cross-classification, interaction terms are easier to interpret than the main effects (Bishop, Fienberg, and Holland, 1975, p. 34). Thus it makes sense to examine cohort effects in the form of interactions in an age by period table. Similarly, when we measure period and age effects in terms of interaction contrasts in Tables 2 and 3, respectively, we take the other two sets of effects as marginal to the ones of interest.

The Identification Problem

Because cohort is uniquely determined by age group and period, we must determine whether all of the age, period and cohort parameters are estimable, and if they are not all estimable, we must determine which ones are. In the additive model for quantitative response variables analogous to expressions (2)-(3), it is well-known (Mason et al., 1973) that all of the parameters are not estimable. The same is the case for the logistic response model.

For the 3-age and 3-period situation we might think we have, in model (2)-(3), 2 independent parameters for age, 2 for period, and 4 for cohort. Since there are 9 expected log-odds, $\{\Omega_{ijk}\}$, under the model and 8 independent parameters plus one for the constant we might expect to be able to estimate all of the parameters. To determine whether this is so we arrange the $\{\Omega_{ijk}\}$ in three different ways. First, we place them in an age by period array (corresponding to the layout of Table 1) and find that there appear to be 4 nonredundant interaction contrasts, involving 2×2 subtables of adjacent cells, for estimating cohort contrasts. Second, we arrange the $\{\Omega_{ijk}\}$ in an age by cohort array (corresponding to the layout of Table 2) and find that there appear to be 2 nonredundant interaction contrasts for estimating period effects. Third, we arrange the $\{\Omega_{ijk}\}$ in a period by cohort array (corresponding to the layout of Table 3) and find that there appear to be 2 nonredundant interaction contrasts for estimating age effects. But these superficial counts of interaction contrasts are deceiving, as we now show.

There are 2 interaction contrasts for the period by cohort array. The first involves periods 1 and 2 crossed with cohorts 3 and 4, and the second involves periods 2 and 3 crossed with cohorts 2 and 3:

$$\Omega_{113} - \Omega_{214} - \Omega_{223} + \Omega_{324} = W_{1(1)} - 2W_{1(2)} + W_{1(3)} = -3W_{1(2)} \quad (7)$$

(since $W_{1(1)} + W_{1(2)} + W_{1(3)} = 0$), and

$$\Omega_{122} - \Omega_{223} - \Omega_{232} + \Omega_{333} = W_{1(1)} - 2W_{1(2)} + W_{1(3)} = -3W_{1(2)}. \quad (8)$$

The two interaction contrasts for the log-odds-ratios turn out to be the same contrast for the age effect parameters, $W_{1(2)}$, and this is the only estimable age effect. Similarly, the two interaction contrasts for 2 × 2 subtables of adjacent cells in the age by cohort array also both reduce to the same contrast for the period effect parameters, $W_{2(2)}$, and this is the only estimable period effect.

Finally, for the age by period array the 4 interaction contrasts involving 2 × 2 subtables of adjacent cells are:

$$\Omega_{113} - \Omega_{122} - \Omega_{214} + \Omega_{223} = 2W_{3(3)} - W_{3(2)} - W_{3(4)} \tag{9}$$

$$\Omega_{122} - \Omega_{131} - \Omega_{223} + \Omega_{232} = 2W_{3(2)} - W_{3(1)} - W_{3(3)} \quad , \tag{10}$$

$$\Omega_{214} - \Omega_{223} - \Omega_{315} + \Omega_{324} = 2W_{3(4)} - W_{3(3)} - W_{3(5)} \quad , \tag{11}$$

and

$$\Omega_{223} - \Omega_{232} - \Omega_{324} + \Omega_{333} = 2W_{3(3)} - W_{3(2)} - W_{3(4)} \quad . \tag{12}$$

Since expressions (9) and (12) involve the same contrast, there are only 3 equations with which to estimate 4 independent cohort parameters. By taking linear combinations of expressions (9), (10) and (11) and using the basic constraint $\Sigma_k W_{3(k)} = 0$, we find the estimable cohort effects to be $W_{3(3)}$, $W_{3(1)} + W_{3(5)}$, and $W_{3(1)} + 2W_{3(4)}$.

Note that we are one parameter short for each type of effect. The difficulty here is the one alluded to in the preceding subsection; namely, interaction contrasts treat the main effects as marginal to the interaction effects. Thus in all three arrays, the interaction-space is of dimension one less than we naively expected. Geometrically, this loss of a single dimension results from one dimension of the interaction space lying completely within the margin subspaces.

Another way to view this identification problem is also illuminating. Since $\sum_i W_{1(i)} = 0$, the estimable parameter $W_{1(2)}$ allows us to estimate the quadratic effect for age, $W_{1(1)} - 2W_{1(2)} + W_{1(3)}$. What we are unable to estimate is the linear effect, $W_{1(3)} - W_{1(1)}$. We noted in the Introduction that a pure linear age effect is indistinguishable from period and cohort effects. In fact what we have seen here is that the linear component of the age effects is indistinguishable from the period and cohort effects. Similarily, the linear components of the period and cohort effects are indistinguishable from the age and cohort effects, and the age and period effects, respectively. Thus the linear components of the effect parameters are the cause of our identification problem, and fitting a model with, say, only linear and quadratic effects does not provide a solution. These comments apply not only in the present context of discrete response variables but also in the case of a continuous response variable as considered by Mason et al. (1973).

The identification problem complicates analysis, but is in many situations surmountable. In this connection there are a number of relevant considerations. First, the fit of an age-period-cohort model to data can be ascertained despite the inestimability of certain of the parameters, as is shown below. Second, by imposing an <u>identification specification</u>, that is, a restriction on a subset of the parameters, the remainder becomes estimable. As we show below, a restriction such as $W_{1(1)}$ = constant, or $W_{1(1)} = W_{1(2)}$, or $W_{2(1)} = W_{2(2)}$, or $W_{3(1)} = W_{3(2)}$ suffices to identify all parameters. Moreover, it turns out that if only a single such restriction is made, then the identification specification is <u>just</u>-identifying and the estimated expected frequencies or log-odds are affected neither by the restriction nor its placement. Third, identification specifications can be <u>over</u>-identifying, and are so if they contain more restrictions than needed to make all parameters estimable. Models with over-identifying specifications can fit data no better than just-identified or under-identified models, and different over-identified models may fit the data differently. Thus, alternative over-identified specifications lead to effectively different models. It is generally possible to make over-identifying restrictions on the basis of prior knowledge and reasoning. We illustrate the use of over-identifying restrictions in the analysis of the empirical example.

Estimated Expected Cell Values

Given the logistic response model (2)-(3), and 3 independent simple random samples at 3 properly spaced points in time, we can compute maximum likelihood estimates of the expected values, $\{m_{ijk\ell}\}$, corresponding to the observed counts in Table 1 using the general results for loglinear models (Fienberg, 1977; Haberman, 1974). These general results tell us that the minimal sufficient statistics for model (1)-(2) are given by the "marginal" totals

$$\{x_{ijk+}\}, \quad \{x_{i++\ell}\}, \quad \{x_{+j+\ell}\}, \quad \{x_{++k\ell}\} \quad , \tag{13}$$

(where $x_{++k\ell}$ is the summation of $x_{ijk\ell}$ over all i and j such that $i - j = k - 3$). We include the totals $\{x_{ijk+}\}$ among the minimal sufficient statistics even though they are formally treated as fixed by the analysis because when the logistic response model is viewed as a special case of a loglinear model these totals are not necessarily treated as fixed. The general theory next tells us that the likelihood equations are found by setting these minimal sufficient statistics equal to their expected values:

$$\hat{m}_{ijk+} = x_{ijk+}\ , \quad i = 1,\ 2,\ 3;\ j = 1,\ 2,\ 3;\ (i - j = k - 3) \tag{14}$$

$$\hat{m}_{i++\ell} = x_{i++\ell}\ , \quad i = 1,\ 2,\ 3;\ \ell = 1,\ 2; \tag{15}$$

$$\hat{m}_{+j+\ell} = x_{+j+\ell}\ , \quad j = 1,\ 2,\ 3;\ \ell = 1,\ 2; \tag{16}$$

$$\hat{m}_{++k\ell} = x_{++k\ell}\ , \quad k = 1,\ 2, \ldots,\ 5;\ \ell = 1,\ 2. \tag{17}$$

Note that equation (17) implies that for cohorts 1 and 5,

$$\hat{m}_{131\ell} = x_{131\ell} \text{ and } \hat{m}_{315\ell} = x_{315\ell} \tag{18}$$

for $\ell = 1, 2$. Furthermore, some algebraic manipulation $(\hat{m}_{1++\ell} + \hat{m}_{2++\ell} - \hat{m}_{+1+\ell} + \hat{m}_{131\ell} - \hat{m}_{++4\ell})$ leads to

$$\hat{m}_{223\ell} = x_{223\ell} \tag{19}$$

for $\ell = 1, 2$. Expression (19) is particular to the 3-age and 3-period case, whereas expression (18) is true in general. Both expressions will be used below, in the subsection on degrees of freedom and goodness of fit.

Alternatives to maximum likelihood estimation are available for this problem, and others might choose to use the weighted least squares or minimum modified chi-square approach of Grizzle, Starmer, and Koch (1969), or generalized (iteratively reweighted) least squares (GLS). For the class of problems we consider here, GLS yields maximum likelihood estimates, and is equivalent to the method of scoring or the Newton-Raphson method of solving the likelihood equations.

In the following subsections we discuss two methods for solving equations (14)-(17): iterative proportional fitting and Newton-Raphson.

Iterative Proportional Fitting

The likelihood equations are solved using the method of iterative proportional fitting, starting with initial values, $\hat{m}^{(0)}_{ijk\ell} = 1$, and successively making multiplicative adjustments to ensure that equations (14), (15), (16), and (17), in turn, hold exactly. Thus, for the ν-th cycle of the iteration ($\nu \geq 0$) we compute

$$\hat{m}^{(4\nu+1)}_{ijk\ell} = \hat{m}^{(4\nu)}_{ijk\ell} \cdot \left(\frac{x_{ijk+}}{\hat{m}^{(4\nu)}_{ijk+}} \right), \tag{20}$$

$$\hat{m}^{(4\nu+2)}_{ijk\ell} = \hat{m}^{(4\nu+1)}_{ijk\ell} \cdot \left(\frac{x_{i++\ell}}{\hat{m}^{(4\nu+1)}_{i++\ell}} \right), \tag{21}$$

$$\hat{m}^{(4\nu+3)}_{ijk\ell} = \hat{m}^{(4\nu+2)}_{ijk\ell} \cdot \left(\frac{x_{+j+\ell}}{\hat{m}^{(4\nu+2)}_{+j+\ell}} \right), \tag{22}$$

$$\hat{m}^{(4\nu+4)}_{ijk\ell} = \hat{m}^{(4\nu+3)}_{ijk\ell} \cdot \left(\frac{x_{++k\ell}}{\hat{m}^{(4\nu+3)}_{++k\ell}} \right), \tag{23}$$

for all i, j, k, and ℓ. This procedure converges to the MLEs, $\{m_{ijk\ell}\}$, and is based on the same idea as the use of iterative proportional fitting for the model of quasi-independence in an unfolded three-dimensional table (with diagonals representing the third dimension) as described in Bishop, Fienberg, and Holland (1975, pp. 225-8) in the context of multiplicative models for social mobility tables. Following their prescription we convert the three-dimensional array in Table 1 into the four-dimensional array of Table 4 with the added dimension corresponding to the third subscript for cohorts. (This four-dimensional array contains a large number of structural zeros denoted by dashes in the table.) Unfolded tables not only have a conceptual and pedagogical value, but they also allow use of standard iterative proportional fitting computer programs (Fay and Goodman, 1973; Haberman, 1973) to estimate the expected frequencies of the age-period-cohort model for given data. What makes the standard computer programs applicable here is that the totals in expression (13) become the marginal totals for the unfolded data in Table 4.

Table 4 here

The use of iterative proportional fitting as outlined in expressions (20)-(23) is identical to the fitting of a loglinear model to the incomplete unfolded four-dimensional array. For under- or just-identified age-period-cohort models one fits the [APC], [AR], [PR] and [CR] configuration, where A, P, C and R denote age, period, cohort and the response variable, respectively.[5] For over-identified age-period-cohort models it is necessary to partially collapse the unfolded array in accordance with the identifying restrictions. If these are, for example, $W_{1(1)} = W_{1(2)}$ and $W_{2(1)} = W_{2(2)}$, then one combines the first 2 rows and first 2 columns in each layer of Table 4. Although this does not happen for all specifications, in this instance collapsing the first 2 rows and 2 columns of each layer of Table 4 combines the frequencies x_{1131} and x_{2231} and combines x_{1132} and x_{2232}. Iterative proportional fitting of the age-period-cohort model for the partially collapsed array then gives $\hat{M}_1 = (x_{1131} + x_{2231})$ and $\hat{M}_2 = (x_{1132} + x_{2232})$, whereas what we would like is $\hat{m}_{1131}$, $\hat{m}_{2231}$, $\hat{m}_{1132}$ and $\hat{m}_{2232}$. To obtain these four estimates we note that according to the model,

$$\hat{M}_1/(\hat{M}_1 + \hat{M}_2) = \hat{m}_{1131}/(\hat{m}_{1131} + \hat{m}_{1132}) = \hat{m}_{2231}/(\hat{m}_{2231} + \hat{m}_{2232})$$

with conditions

$$x_{1131} + x_{1132} = \hat{m}_{1131} + \hat{m}_{1132} \text{ and } x_{2231} + x_{2232} = \hat{m}_{2231} + \hat{m}_{2232} .$$

Since we know all observed frequencies and $\hat{M}_1$ and $\hat{M}_2$, we can determine $\hat{m}_{1131}$, $\hat{m}_{1132}$, $\hat{m}_{2231}$ and $\hat{m}_{2232}$. Having secured these four estimated expected frequencies we expand the estimated array back to the original unfolded dimensions, inserting the four estimated expected frequencies into the correct cells. It is important to be aware that there is actually one more degree of freedom than appears (spuriously) to be the case from the partially collapsed table, and that assessment of goodness-of-fit from the partially collapsed table gives an incorrect answer. The array must be re-expanded. Thus, when nonstructural zero cells are lost because the identifying restrictions require the combining of nonstructural zero cells for iterative proportional fitting, a certain amount of additional calculation is necessary. Generalized iterative proportional fitting (Darroch and Ratcliffe, 1972) requires neither unfolding the table nor combining of frequencies, and allows different kinds of restrictions on parameters than iterative proportional fitting does. The generalized algorithm is not yet routinely included in most general purpose programs.

An additional problem with the iterative proportional fitting algorithm is that it can be slow to converge for unfolded tables. For example, using artificial data for the $4 \times 4 \times 2$ case, unfolding the array and fitting just-identified age-period-cohort models required as few as 6 iterations and as many as 69 (using for all specifications the same criterion to terminate iterations), depending on the placement of the identifying specification. This slowness of convergence is a serious disadvantage of iterative proportional fitting, since not only unnecessary expense is incurred, but also undue rounding errors can be encountered. On the other hand, iterative proportional fitting does have the advantage that it allows the fitting of under- and just-identified age-period-cohort models and the assessment of their fit without our directly dealing with effect parameter estimates.

Newton-Raphson Procedure

Newton-Raphson and iterative proportional fitting are alternative methods of solving the likelihood equations. They differ in that iterative proportional fitting estimates cell frequencies from which parameter estimates are then derived, and Newton-Raphson estimates parameters from which estimated expected frequencies are then derived. The advantages of Newton-Raphson are that it converges more rapidly and that it produces estimated variances for the parameters in the model as a byproduct. Since the Newton-Raphson procedure estimates parameters rather than expected frequencies, its use requires prior resolution of the identification problem. In addition, the arrangement of the data for Newton-Raphson computations differs from that of iterative proportional fitting and resembles the arrangement for regression problems. Hence linear model treatments of identifiability carry over to the discrete context when the data are arranged for Newton-Raphson computations. The remainder of this section illustrates the identification problem and its resolution for the 3 × 3 × 2 case when the data are set up in a regression format.

We begin by rewriting expression (2) in the regression form

$$\Omega_{ijk} = \alpha + \Sigma_i \beta_i A_i + \Sigma_j \gamma_j P_j + \Sigma_k \delta_k C_k \, , \tag{24}$$

where the A_i, P_j, and C_k are dummy variables denoting the i-th of 3 age groups, the j-th of 3 periods and the k-th of 5 cohorts, respectively. The β_i, γ_j and δ_k are the parameters for age, period and cohort respectively, and α is a constant. The Ω_{ijk} retain their meaning as logits expected under the model. Expression (24) can be stated in matrix form as $\underset{\sim}{\Omega} = \underset{\sim}{X}\underset{\sim}{\beta}$ (where "~" denotes a vector or matrix), or

$$
\underbrace{\begin{pmatrix} \Omega_{113} \\ \Omega_{123} \\ \Omega_{131} \\ \Omega_{214} \\ \Omega_{223} \\ \Omega_{232} \\ \Omega_{315} \\ \Omega_{324} \\ \Omega_{333} \end{pmatrix}}_{\underset{\sim}{\Omega}}
=
\underbrace{\begin{array}{c}
\begin{array}{cccccccccccc}
\underset{\sim}{1} & \underset{\sim}{A}_1 & \underset{\sim}{A}_2 & \underset{\sim}{A}_3 & \underset{\sim}{P}_1 & \underset{\sim}{P}_2 & \underset{\sim}{P}_3 & \underset{\sim}{C}_1 & \underset{\sim}{C}_2 & \underset{\sim}{C}_3 & \underset{\sim}{C}_4 & \underset{\sim}{C}_5
\end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\end{array}}_{\underset{\sim}{X}}
\underbrace{\begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{pmatrix}}_{\underset{\sim}{\beta}}, \qquad (25)
$$

where $\underset{\sim}{\Omega}$ is the vector of logits expected under the model, $\underset{\sim}{X}$ is the design matrix comprised of vectors of dummy variables (with age, period and cohort category labels at the top) and a vector of ones ($\underset{\sim}{\iota}$) for the constant, and $\underset{\sim}{\beta}$ is the vector of effect parameters.

Expression (24) requires a normalization, such as (3) which constrains the effects of each classification to sum to zero, for two reasons: First, there are 9 independent logits under the model, but the design matrix $\underset{\sim}{X}$ contains 12 columns, and thus the parameters of $\underset{\sim}{\beta}$ are not identified. Second, and more importantly for larger dimension problems, $\underset{\sim}{X}$ is not of full column rank, since

$$\underset{\sim}{\iota} = \underset{\sim}{A}_1 + \underset{\sim}{A}_2 + \underset{\sim}{A}_3 = \underset{\sim}{P}_1 + \underset{\sim}{P}_2 + \underset{\sim}{P}_3 = \underset{\sim}{C}_1 + \underset{\sim}{C}_2 + \underset{\sim}{C}_3 + \underset{\sim}{C}_4 + \underset{\sim}{C}_5 \quad ,$$

which implies that $\underset{\sim}{X}'\underset{\sim}{X}$ is not invertible. Estimation of the parameters by Newton-Raphson requires a full column rank matrix.

One normalization in which the sums of the coefficients of age, period and cohort are zero is based on the transformed design matrix,

$$\underset{\sim}{X}^* = (\underset{\sim}{\iota}, \underset{\sim}{A}_1 - \underset{\sim}{A}_3, \underset{\sim}{A}_2 - \underset{\sim}{A}_3, \underset{\sim}{P}_1 - \underset{\sim}{P}_3, \underset{\sim}{P}_2 - \underset{\sim}{P}_3, \underset{\sim}{C}_1 - \underset{\sim}{C}_5, \underset{\sim}{C}_2 - \underset{\sim}{C}_5,$$

$$\underset{\sim}{C}_3 - \underset{\sim}{C}_5, \underset{\sim}{C}_4 - \underset{\sim}{C}_5) \quad . \tag{26}$$

If $\underset{\sim}{\beta}^*$ is the (column) vector of coefficients associated with $\underset{\sim}{X}^*$, it follows that

$$\underset{\sim}{\beta}^{*\prime} = (\alpha^*, \beta_1^* = \beta_1 - \beta_3, \beta_2^* = \beta_2 - \beta_3, \gamma_1^* = \gamma_1 - \gamma_3, \gamma_2^* = \gamma_2 - \gamma_3,$$

$$\delta_1^* = \delta_1 - \delta_5, \delta_2^* = \delta_2 - \delta_5, \delta_3^* = \delta_3 - \delta_5, \delta_4^* = \delta_4 - \delta_5) \quad ,$$

and we observe that for the age effects

$$\beta_1^*(\underset{\sim}{A}_1 - \underset{\sim}{A}_3) + \beta_2^*(\underset{\sim}{A}_2 - \underset{\sim}{A}_3) = \beta_1^*\underset{\sim}{A}_1 + \beta_2^*\underset{\sim}{A}_2 - (\beta_1^* + \beta_2^*)\underset{\sim}{A}_3,$$

and so on for the period and cohort effects. Writing

$$\begin{aligned} \beta_3^* &= -(\beta_1^* + \beta_2^*), \\ \gamma_3^* &= -(\gamma_1^* + \gamma_2^*), \\ \delta_5^* &= -(\delta_1^* + \delta_2^* + \delta_3^* + \delta_4^*), \end{aligned} \tag{27}$$

it follows that $\underset{\sim}{\beta}^*$ is a set of contrasts based on $\underset{\sim}{\beta}$ and that together with β_3^*, γ_3^* and δ_5^* the contrasts induced by (26) are symmetric, as in (3). With this normalization the model becomes

$$\underset{\sim}{\Omega} = \underset{\sim}{X}^* \underset{\sim}{\beta}^* \quad , \tag{28}$$

with β_3^*, γ_3^* and δ_5^* defined in (27).

Imposition of a coefficient normalization does not solve the identification problem. The discussion of Mason et al. (1973) applies here, but it is also useful to reinforce the insight that it is the <u>linear</u> effects of age, period and cohort which are not estimable. This point was made in an earlier section and also by Winsborough (1976). That age, period and cohort categories are scalable with respect to time provides the basis for a simple explanation of the locus of the identification problem. In the 3 × 3 × 2 case, the age, period and cohort model can be represented by

$$\Omega_{ijk} = \xi_0 + \xi_1 A + \xi_2 A^2 + \xi_3 P + \xi_4 P^2 + \xi_5 C + \xi_6 C^2 + \xi_7 C^3 + \xi_8 C^4 \quad , \tag{29}$$

where A is age, P is period and C is cohort, all variables are scaled with respect to time (e.g., years), and the highest powers of age, period and cohort in (29) are exactly one less than the number of ages, periods and cohorts, respectively. In general, the effects of an I-category dimension in a (log)linear model can always be represented by a polynomial of degree I-1. Thus expressions (28) and (29) are equivalent. Therefore, some of the coefficients of (29), in particular those of the linear terms, are not identified. Substitution of $A = P - C$ into (29) gives

$$\Omega_{ijk} = \alpha + (\xi_1 + \xi_3)P + (\xi_2 + \xi_4)P^2 + (\xi_5 - \xi_1)C + (\xi_2 + \xi_6)C^2$$

$$+ \xi_7 C^3 + \xi_8 C^4 - 2\xi_2 PC \quad , \tag{30}$$

and it is clear that all coefficients except ξ_1, ξ_3 and ξ_5 are estimable. Hence the linear effects of age, period and cohort are indistinguishable and all of the higher order effects of these dimensions are distinguishable.

Estimation of all coefficients in model (28), or equivalently model (29), requires an identification specification--a restriction which eliminates the linear dependence of age, period and cohort. Linear restrictions are simple and may often be appropriate. It may be most comformable with data and substantive theory and knowledge to equate the effects of two (or more) adjacent ages, periods or cohorts. For example, in the 3 × 3 × 2 case, one identification restriction is $\beta_1^* = \beta_2^*$. Imposing this restriction on model (28) gives a design matrix which can be written as

$$\underset{\sim}{L} = (\underset{\sim}{\iota}, \underset{\sim}{A}_1 + \underset{\sim}{A}_2 - 2\underset{\sim}{A}_3, \underset{\sim}{P}_1 - \underset{\sim}{P}_3, \underset{\sim}{P}_2 - \underset{\sim}{P}_3, \underset{\sim}{C}_1 - \underset{\sim}{C}_5, \underset{\sim}{C}_2 - \underset{\sim}{C}_5, \underset{\sim}{C}_3 - \underset{\sim}{C}_5,$$

$$\underset{\sim}{C}_4 - \underset{\sim}{C}_5) \tag{31}$$

which has full column rank, and has an associated vector of coefficients

$$\underset{\sim}{B}' = (a, b_1, g_1, g_2, d_1, d_2, d_3, d_4) \quad . \tag{32}$$

Here a is the constant, and the rest of the coefficients are contrasts of the original $\underset{\sim}{\beta}$. In particular,

$$b_1 = \beta_1 + \beta_2 - 2\beta_3 = 2\beta_1 - 2\beta_3 = 2\beta_2 - 2\beta_3$$

according to the identifying restriction, and

$$g_1 = \gamma_1 - \gamma_3,\ g_2 = \gamma_2 - \gamma_3,$$

$$d_1 = \delta_1 - \delta_5,\ d_2 = \delta_2 - \delta_5,\ d_3 = \delta_3 - \delta_5,\ d_4 = \delta_4 - \delta_5\ ,$$

and because of the structure of $\underset{\sim}{L}$,

$$b_1 = b_2,\ b_3 = -2b_1\ ,$$

$$g_3 = -(g_1 + g_2)\ ,$$

$$d_5 = -(d_1 + d_2 + d_3 + d_4)\ ,$$

so that the coefficients are still symmetrically normalized.

If maximum likelihood estimates exist, the expression

$$\underset{\sim}{\hat{\Omega}} = \underset{\sim}{L}\underset{\sim}{\hat{B}} \tag{33}$$

can be arrived at by the Newton-Raphson procedure, where $\underset{\sim}{\hat{\Omega}}$ is the vector of estimated logits expected under the model, $\underset{\sim}{L}$ is as defined in (31) and $\underset{\sim}{\hat{B}}$ is an estimate of $\underset{\sim}{B}$ defined in (32). We note that

$$\underset{\sim}{\hat{B}} = (\underset{\sim}{L}'\underset{\sim}{L})^{-1}\underset{\sim}{L}'\underset{\sim}{\hat{\Omega}}\ , \tag{34}$$

which indicates the reason for requiring $\underset{\sim}{L}$ to be full column rank. The vector of coefficients $\hat{\underset{\sim}{B}}$ is arrived at by a process which is in fact a series of convergent regressions. This point is described in detail by Haberman (1978). Additional discussion of the Newton-Raphson procedure is available in Haberman (1974) and Bock (1975). The procedure has been programmed for discrete data analysis by Bock and Yates (1973).

Degrees of Freedom and Goodness-of-Fit

From the discussion of identification in the $3 \times 3 \times 2$ case using interaction contrasts we know that the total number of independent parameters corresponding to the age, period and cohort effects to be estimated from the logistic response model (2)-(3) is 7 rather than 8. Since there are 9 odds-ratios and 7 independent effect parameters plus a constant, W, we have one degree of freedom associated with the full model, and thus we can assess its goodness-of-fit.

The use of iterative proportional fitting or Newton-Raphson in this special case of a $3 \times 3 \times 2$ table is equivalent to finding a unique quantity Δ satisfying certain constraints. To see this we first remove the $(1,3,1,\ell)$, $(3,1,5,\ell)$ and $(2,2,3,\ell)$ cells for $\ell = 1,2$ from Table 1 since they are fitted exactly by maximum likelihood (see expressions (18) and (19)). Within each layer the likelihood equations for rows reduce to

$$\hat{m}_{113\ell} + \hat{m}_{122\ell} = x_{113\ell} + x_{122\ell} , \tag{35}$$

$$\hat{m}_{214\ell} + \hat{m}_{232\ell} = x_{214\ell} + x_{232\ell} , \tag{36}$$

$$\hat{m}_{324\ell} + \hat{m}_{333\ell} = x_{324\ell} + x_{333\ell} , \tag{37}$$

for columns they reduce to

$$\hat{m}_{113\ell} + \hat{m}_{214\ell} = x_{113\ell} + x_{214\ell} , \tag{38}$$

$$\hat{m}_{122\ell} + \hat{m}_{324\ell} = x_{122\ell} + x_{324\ell} , \tag{39}$$

$$\hat{m}_{213\ell} + \hat{m}_{333\ell} = x_{231\ell} + x_{333\ell} , \tag{40}$$

and for diagonals they reduce to

$$\hat{m}_{122\ell} + \hat{m}_{232\ell} = x_{122\ell} + x_{232\ell} , \tag{41}$$

$$\hat{m}_{113\ell} + \hat{m}_{333\ell} = x_{113\ell} + x_{333\ell} , \tag{42}$$

$$\hat{m}_{214\ell} + \hat{m}_{324\ell} = x_{214\ell} + x_{324\ell} . \tag{43}$$

Now suppose we express $\hat{m}_{1131}$ as

$$\hat{m}_{1131} = \Delta + x_{1131} .$$

Then expressions (35), (38) and (42) imply that

$$\hat{m}_{1221} = -\Delta + x_{1221} ,$$

$$\hat{m}_{2141} = -\Delta + x_{2142} ,$$

$$\hat{m}_{3331} = -\Delta + x_{3331} ,$$

and expressions (41) and (43) imply that

$$\hat{m}_{2321} = \Delta + x_{2321} ,$$

$$\hat{m}_{3241} = \Delta + x_{3241} .$$

Equation (14) then yields similar expressions for the negative response layer with Δ replaced by $-\Delta$ everywhere. Thus, solving the likelihood equations in this special case is equivalent to finding a unique quantity Δ such that by adding the entries in Table 5 to those in Table 1 we get positive expected values satisfying the logistic response model (2)-(3). The quantity Δ corresponds to the single degree of freedom associated with the model. (The approach taken here is similar to that of Bartlett (1935) for the no-second-order interaction model in a $2 \times 2 \times 2$ table. In Bartlett's case and ours there is a single degree of freedom and a single quantity Δ to be computed.)

Table 5 here

The MLEs, $\{\hat{m}_{ijk\ell}\}$, found by solving the likelihood equations will all be strictly positive if all the observed counts are positive, or if there is exactly one observed zero count (with the 6 counts $x_{131\ell}$, $x_{315\ell}$, and $x_{223\ell}$ ($\ell = 1, 2$) all positive). When two or more such observed counts are zero they must correspond to cells in Table 5 having the same "sign" on the Δ; otherwise one will have estimated expected value Δ and the other $-\Delta$. Since expected values are nonnegative this implies that $\Delta = 0$, that the estimated expected values will be identical with the observed, and that there will be zero degrees of freedom.

Sporadic zero cell counts are less problematic when there are more age groups and more periods. With respect to zero cell counts in the special case considered here our analysis is again closely related to corresponding considerations for the model of no-second-order interaction in a $2 \times 2 \times 2$ table (Haberman, 1976; Fienberg, 1977).

Having computed the estimated expected cell values, we can test the goodness-of-fit of model (2)-(3) using either the Pearson statistic,

$$X^2 = \Sigma \frac{(x_{ijk\ell} - \hat{m}_{ijk\ell})^2}{\hat{m}_{ijk\ell}} \tag{44}$$

or the likelihood ratio statistic,

$$G^2 = 2 \, \Sigma \, x_{ijk\ell} \log \frac{x_{ijk\ell}}{\hat{m}_{ijk\ell}} \tag{45}$$

If the model is correct then either statistic is asymptotically distributed as a chi-square variate with one degree of freedom. If the model is incorrect the statistics have asymptotic non-central chi-square distributions that are stochastically larger than the corresponding null distribution.

Computing Effect Parameters Directly From Expected Values

Given estimated expected cell values obtained by iterative proportional fitting, it is possible to estimate the effect parameters. The difficulty in these computations stems from the nonorthogonality of the components in the design matrix associated with the logistic response model, (2)-(3). In the 3-age, 3-period and 5-cohort situation the computation of estimated effect parameters is relatively simple, but nevertheless it is dependent on the identification specification chosen. From our above investigation of which parameters are estimable, if

$$\hat{\Omega}_{ijk} = \log\left(\frac{\hat{m}_{ijk1}}{\hat{m}_{ijk2}}\right) , \tag{46}$$

then

$$\hat{W}_{1(2)} = (\hat{\Omega}_{214} + \hat{\Omega}_{223} - \hat{\Omega}_{113} - \hat{\Omega}_{324})/3 , \tag{47}$$

$$\hat{W}_{2(2)} = (\hat{\Omega}_{122} + \hat{\Omega}_{223} - \hat{\Omega}_{113} - \hat{\Omega}_{232})/3 , \tag{48}$$

and

$$\begin{aligned} \hat{W}_{3(3)} = {} & [(\hat{\Omega}_{113} + \hat{\Omega}_{223} - \hat{\Omega}_{214} - \hat{\Omega}_{122}) \\ & + (\hat{\Omega}_{113} + \hat{\Omega}_{333} - \hat{\Omega}_{131} - \hat{\Omega}_{315})]/5 , \end{aligned} \tag{49}$$

$$\begin{aligned} \hat{W}_{3(1)} + \hat{W}_{3(5)} = {} & (\hat{\Omega}_{113} + \hat{\Omega}_{223} - \hat{\Omega}_{214} - \hat{\Omega}_{122})(2/5) \\ & - (\hat{\Omega}_{113} + \hat{\Omega}_{333} - \hat{\Omega}_{131} - \hat{\Omega}_{315})(3/5) , \end{aligned} \tag{50}$$

$$\hat{W}_{3(1)} + 2\hat{W}_{3(4)} = (\hat{\Omega}_{113} + \hat{\Omega}_{324} - \hat{\Omega}_{122} - \hat{\Omega}_{315})$$

$$- (\hat{\Omega}_{113} + \hat{\Omega}_{223} - \hat{\Omega}_{214} - \hat{\Omega}_{122})(2/5)$$

$$- (\hat{\Omega}_{113} + \hat{\Omega}_{333} - \hat{\Omega}_{131} - \hat{\Omega}_{315})(2/5) \ . \quad (51)$$

As noted earlier, these estimates involve 2 × 2 tables of log-odds-ratios, and the constraints given in expression (3). If the identification specification is just-identifying, and involves only the age parameters, for example, then $\hat{W}_{1(1)}$ and $\hat{W}_{1(3)}$ are directly computable (since $W_{1(1)} + W_{1(2)} + W_{1(3)} = 0$). Having estimated the age effects we can estimate period and cohort effects by using the differences for pairs of log-odds-ratios in the same cohort and then in the same period, respectively, since

$$\Omega_{ijk} - \Omega_{i'j'k} = W_{1(i)} - W_{1(i')} + W_{2(j)} - W_{2(j')} \quad (52)$$

(for $i \neq i'$ and $i - i' = j - j'$), and

$$\Omega_{ijk} - \Omega_{i'jk'} = W_{1(i)} - W_{1(i')} + W_{3(k)} - W_{3(k')} \quad (53)$$

(for $i \neq i'$ and $i - i' = k - k'$) .

If an over-identifying restriction is made, then estimated expected frequencies are obtained as described earlier, and the calculation of the estimated effect parameters follows along the lines just described, modified only by the simplifications introduced by the additional restriction(s).

Reduced Models

If the full logistic response model with age, period and cohort effects, whether under-identified, just-identified or over-identified, provides an acceptable fit to the data, then it will usually be of interest to explore whether we can set the effects of one or two of these dimensions to zero. Fitting age-period, age-cohort, cohort-period models, and even further reduced models, is a straightforward task with any computer program designed to fit standard loglinear models to multidimensional arrays. In particular, such reduced models pose no special identification problems because there is no way for the linear component of one type of effect to become confounded with the linear components of the other two types.

To fit the model with only age and period effects, using iterative proportional fitting, we compute estimated expected cell values for Table 1 (ignoring the cohort subscript) under the model of no-second-order interaction, basing the calculations on equations (14)-(16). With Newton-Raphson, using, say, the normalized design matrix (26), one simply omits the columns pertaining to cohort contrasts, and estimation proceeds as for the age-period-cohort model. The age-period model has 4 degrees of freedom.

To fit the age-cohort model, using iterative proportional fitting, we compute estimated expected cell values for the <u>incomplete</u> array, one layer of which is given in Table 2. Again we use the model of no-second-order interaction, this time basing the calculations on equations (14), (15) and (17), ignoring the period subscript, and using initial values of zero in the cells with structural zeros. With Newton-Raphson, again using the normalized design matrix (26), one omits the columns pertaining to the period contrasts and estimation proceeds as before. This model has 2 degrees of freedom. Similar comments apply to fitting the period-cohort model, for which the estimated expected cell values must satisfy equations (14), (16) and (17). The period-cohort model also has 2 degrees of freedom.

In addition we can fit reduced models with only one set of effect parameters. Table 1 arrays the frequencies conveniently for iterative proportional fitting of the age or period models. Tables 2 or 3 indicate convenient forms of data array for iterative proportional fitting of the cohort model. With Newton-Raphson, the normalized design matrix (26) will exclude the period and cohort columns, the age and cohort columns, and the age and period columns for estimation of, respectively, the age, period and cohort models.[6]

If iterative proportional fitting is used, then for any full or reduced model we can compute the estimated effect parameters using the formulae given earlier for the age-period-cohort model. With Newton-Raphson, the effect parameters are estimated directly.

The fit of the reduced models can be assessed using the standard goodness-of-fit statistics, (44) and (45), and can be compared to the fit of the full model using the log-likelihood-ratio statistics for nested models, i.e., the conditional likelihood ratio test for the fit of the reduced model given that the full model is correct. We advocate the use of both the unconditional and the conditional goodness-of-fit statistics since both the reduced model and the full model can fit the data moderately well, but the conditional fit of the reduced model given the full one may be statistically significant at some prechosen level, e.g., 0.01 or 0.05.

I AGE GROUPS, J PERIODS AND A DICHOTOMOUS RESPONSE VARIABLE

The results for the $3 \times 3 \times 2$ case generalize directly to the I-age and J-period situation for which model (2)-(3) was defined. Corresponding to the basic data arrays in Tables 1-3 there are three arrays: age by period by response, of dimension $I \times J \times 2$; age by cohort by response, of dimension $I \times (I + J - 1) \times 2$; period by cohort by response, of dimension $J \times (I + J - 1) \times 2$. There are structural zeros in the age-cohort-response and period-cohort-response arrays for those age-cohort and period-cohort combinations which are logically impossible given the original structure of the data in the age-period-response array.

The Identification Problem

The full response model of (2)-(3) would seem to suggest that there are $I - 1$ independent age effects, $J - 1$ independent period effects and $I + J - 2$ independent cohort effects, for a total of $2(I + J - 2)$. Just as in the 3-age and 3-period situation, however, the linear component of any one set of effect parameters cannot be separated from the linear components of the other two sets of parameters. Thus there are actually $2(I + J - 3)$ independent estimable effect parameters, rather than the $2I + 2J - 3$ that appear in the logistic response model, and the identification problem is the same as in the $3 \times 3 \times 2$ case.

We can estimate $I - 2$ linear contrasts involving the age effect parameters,

$$W_{1(i)} - 2W_{1(i+1)} + W_{1(i+2)} \qquad i = 1,2,\ldots,I-2, \qquad (54)$$

by looking at interaction contrasts in 2 × 2 subtables from the J × K (period by cohort) array of log-odds-ratios, $\{\Omega_{ijk}\}$. Similarly we can estimate J − 2 linear contrasts involving period,

$$W_{3(j)} - 2W_{2(j+1)} + W_{3(j+2)} \qquad j = 1,2,\ldots,J-2, \qquad (55)$$

by examining interaction contrasts in 2 × 2 subtables from the I × K (age by cohort) array of log-odds-ratios, and I + J − 3 linear contrasts involving cohort,

$$W_{3(k)} - 2W_{3(k+1)} + W_{3(k+2)} \qquad k = 1,2,\ldots,I+J-3, \qquad (56)$$

from 2 × 2 subtables of the I × J (age by period) array.

Expressions (54)-(56) measure local quadratic effects in the corresponding subscripts, and reinforce the notion that we cannot estimate the linear components without an identification specification. As noted earlier, if the specification is just-identifying, e.g., $W_{1(1)} = W_{1(2)}$, then it is just like any other assumption in a statistical model that is not capable of direct verification as part of an analysis. It must be grounded in substantive knowledge relating to the data in question or it must come from observations on and analyses of other data for related phenomena.[7] If an over-identifying specification is made, then, conditional on the just-identifying restriction, it is possible to test this additional constraint. For example, if the just-identifying specification is $W_{1(1)} = W_{1(2)}$, and we over-identify with

$W_{1(1)} = W_{1(2)} = W_{1(3)}$, then comparison of the likelihood ratio statistics for the two estimated models will enable a test of the additional constraint. Over-identifying restrictions require no less defense than just-identifying restrictions.

Estimated Expected Cell Values and Reduced Models

The methods described for the 3-age and 3-period case generalize immediately to the I-age and J-period situation. The minimal sufficient statistics are still given by expression (13) and the likelihood equations for the full age, period, and cohort model are identical to those in equations (14)-(17) except that the subscripts i, j, and k now run from $i = 1,2,\ldots,I$; $j = 1,2,\ldots,J$; and $k = 1,2,\ldots,I+J-1$. We can still use the method of iterative proportional fitting (as described by (20)-(23)) to compute estimated expected cell values, and given an identification specification we can use the estimable contrasts in expressions (54)-(56) along with equations (52)-(53) to solve for the estimated effect parameters. Alternatively, given an identification specification, we can estimate the parameters directly using Newton-Raphson. And, as in the special case of 3-ages and 3-periods, we can calculate the estimated expected values and parameter estimates for reduced models.

The only major changes in going from the 3-age and 3-period case to the general situation involve the degrees of freedom associated with the various models. For example, the full model has $(I - 2)(J - 2)$ degrees of freedom associated with assessing its fit. Table 6 lists the full model and the 7 possible reduced models, and the associated minimal sufficient statistics and degrees of freedom. To read the table for the I × J × 2 case, set H = 1 and L = 2, and determine I and J from the data. Goodman (1975a) gives a special case of this table.

<u>Table 6 here</u>

I AGE GROUPS, J PERIODS AND A POLYTOMOUS RESPONSE VARIABLE

In preceding sections we dealt with logistic models measuring the effects of age, period, and cohort on a dichotomous response variable. Here we consider two extensions of these models for polytomous response variables. Suppose the response variable has L categories, and the basic data arrays of interest are the $I \times J \times L$ age-period-response table, the $I \times (I + J - 1) \times L$ age-cohort-response table, and the $J \times (I + J - 1) \times L$ period-cohort-response table. As before, we label the counts in these arrays using four subscripts, i.e., $\{x_{ijk\ell}\}$ where now $\ell = 1,2,\ldots,L$, and we denote the corresponding expected cell values by $\{m_{ijk\ell}\}$. For dichotomous response variables we considered a model for the single logit structure

$$\log\left(\frac{m_{ijk1}}{m_{ijk2}}\right) .$$

For polytomous response variables we can consider models for L - 1 different logit structures. Two possible (and quite different) ways to define these are:

$$\log\left(\frac{m_{ijk\ell}}{\sum_{\ell'>\ell} m_{ijk\ell'}}\right) \qquad \text{for } \ell = 1,2,\ldots,L-1, \tag{57}$$

and

$$\log\left(\frac{m_{ijk\ell}}{m_{ijk\ell+1}}\right) \qquad \text{for } \ell = 1,2,\ldots,L-1. \tag{58}$$

If we would like to fit models with the same parametric structure to the L - 1 logits, and to have the models correspond as a group to a loglinear model for the $\{m_{ijk\ell}\}$, then our choice would be expression (58). Note that expression (58), in such circumstances, is equivalent to models with the same parametric structure for the L - 1 logits

$$\log\left(\frac{m_{ijk\ell}}{m_{ijkL}}\right) \qquad \text{for } \ell = 1,2,\ldots,L-1, \tag{59}$$

or any other set of L - 1 nonredundant logits for the logarithm of the odds involving pairs of expected values. The generalization of the logistic response model we would consider for expression (58) is

$$\begin{aligned}\Omega_{ijk} &= \log\left(\frac{m_{ijk\ell}}{m_{ijk\overline{\ell+1}}}\right) \\ &= W^{(\ell)} + W^{(\ell)}_{1(i)} + W^{(\ell)}_{2(j)} + W^{(\ell)}_{3(i-j+J)} \quad \text{for } \ell = 1,2,\ldots,L-1,\end{aligned} \tag{60}$$

with

$$\sum_i W^{(\ell)}_{1(i)} = \sum_j W^{(\ell)}_{2(j)} = \sum_k W^{(\ell)}_{3(k)} = 0 \quad . \tag{61}$$

All of the results for the I × J × 2 case carry over immediately to this set of models; however, the summations involving the fourth subscript run up to L instead of 2. Degrees of freedom in Table 6 are read by setting H = 1, with I, J and L determined by the data.

If the L response categories are ordered and it makes substantive sense to think of the effects linking the response variable to age, period, and cohort as increasing linearly with the category number (e.g., the category number represents some latent variable), then we may wish to test for the equality of various effect parameters, e.g.,

$$W_{1(i)}^{(\ell)} = W_{1(i)}^{*} \qquad \text{for } \ell = 1,2,\ldots,L-1. \tag{62}$$

Such reduced models can be handled without trouble using the methodology of loglinear models with ordered categories for some of the variables (Fienberg, 1977).

When the response categories have a natural order, e.g., educational attainment (grade school, high school, college, graduate school), the other choice of logits, in expression (57) may be preferable. The quantities

$$\sum_{\ell' > \ell} m_{ijk\ell'} / m_{ijk\ell}$$

are often referred to as continuation odds, and they are of substantive interest in various fields.[8] Expression (57) gives the negative logarithms of the continuation odds. There is also a technical reason for working with the logits in expression (57). Let $P_{ijk|\ell}$ be the probability of a response in category ℓ given age i and period j, where $\Sigma_\ell P_{ijk|\ell} = 1$. Then, when the $\{x_{ijk\ell}\}$ consist of observations from IJ independent multinomial variates with sample sizes $\{x_{ijk+}\}$ and cell probabilities $\{P_{ijk|\ell}\}$,

$$m_{ijk\ell} = x_{ijk+} P_{ijk|\ell} , \tag{63}$$

so that

$$\frac{m_{ijk\ell}}{\sum\limits_{\ell'>\ell} m_{ijk\ell'}} = \frac{P_{ijk|\ell}}{\sum\limits_{\ell'>\ell} P_{ijk|\ell'}} \quad . \tag{64}$$

We can write the multinomial likelihood functions as products of L − 1 binomial likelihoods, the ℓ-th of which has sample size

$$\{\sum_{\ell'\geq\ell} x_{ijk\ell'}\}$$

and cell probabilities

$$\{P_{ijk|\ell} \Big/ \sum_{\ell'\geq\ell} P_{ijk|\ell'}\} \quad .$$

Then, if we use the method of maximum likelihood to estimate the parameters in the logistic response models

$$\log\left(\frac{m_{ijk\ell}}{\sum\limits_{\ell'>\ell} m_{ijk\ell}}\right) = W^{(\ell)} + W^{(\ell)}_{1(i)} + W^{(\ell)}_{2(j)} + W^{(\ell)}_{3(i-j+J)}, \quad \ell = 1,2,\ldots,L-1, \tag{65}$$

subject to (61), we can estimate each logit model separately using the methods described for the I × J × 2 case, and can simply add individual chi-square statistics to get an overall goodness-of-fit statistic for the set of models. Moreover, the observed binomial proportions

$$x_{ijk\ell} \Big/ \sum_{\ell'\geq\ell} x_{ijk\ell'} \qquad \ell = 1,2,\ldots,L-1, \tag{66}$$

are asymptotically independent of each other so that the fit to the $L - 1$ logit models and various associated reduced models can be assessed independently. In our analysis of the large scale example we use this continuation odds approach for a polytomous response structure.

For the logistic response models in (65) it might be of substantive interest to explore the equality of parameters across models as in expression (62). The estimated expected values for such a class of restricted models and the associated tests of fit can be handled by thinking in terms of a set of counts with 5 subscripts, $\{y_{ijk\ell t}\}$, where

$$y_{ijk\ell t} = \begin{cases} x_{ijk\ell} & \text{for } t = 1 \\ \sum_{\ell' > \ell} x_{ijk\ell'} & \text{for } t = 2 \end{cases} .$$

Now, if we let $m^*_{ijk\ell t}$ be the expected value under model (65) corresponding to $y_{ijk\ell t}$, then we fit the $L - 1$ models simultaneously by fitting a hierarchical loglinear model to the $\{y_{ijk\ell t}\}$ with minimal sufficient statistics:

$$\{y_{ijk+t}\},\quad \{y_{i++\ell t}\},\quad \{y_{+j+\ell t}\},\quad \{y_{++k\ell t}\} .$$

If we restrict the model so that (62) holds we fit the loglinear model with minimal sufficient statistics:

$$\{y_{ijk+t}\}, \quad \{y_{i++\ell+}\}, \quad \{y_{+j+\ell t}\}, \quad \{y_{++k\ell t}\} .$$

Similarly we can handle other reduced models involving equality of period and cohort effects across the L - 1 logistic response structures.

DIFFERENTLY SPACED AGE AND PERIOD INTERVALS

To this point we have discussed the identification and estimation of age-period-cohort models under the assumption that age groups and period intervals are equal and constant in length. Data available for analysis often do not conform to this assumption. For example, many series constructed from U.S. Census data allow relatively detailed age groups, but limit period intervals to decades. As another example, researchers increasingly work with archived sample survey data in which age is coded in great detail (single years), while the surveys are available on a regular but sparser basis (e.g., four-year intervals for presidential election surveys). Moreover, sometimes there are gaps in archived survey series, so that intervals between successive surveys are irregular. In these circumstances it might be possible, and occasionally desirable on substantive grounds, to combine ages and/or periods so that the span of age groups is constant and equals a constant period interval; but this would entail incomplete use of information, might also be <u>un</u>desirable on substantive grounds, and in any case would not always be possible. What should be done? This section discusses the problem of differently spaced age groups and period intervals in two parts. We first address the common and important case in which age group spans are narrower than period intervals, but the spacing is constant and there are $H \geq 2$ contiguous age groups between successive periods. For this case we state the identification problem, indicate how estimation proceeds, and list degrees of freedom. Second, we discuss briefly the case in which period intervals are variable in length. The problems associated with this case have not been completely resolved.

Unequally But Evenly Spaced Age Groups and Period Intervals

The general array we deal with in this case is of size $HI \times J \times L$, since we assume that the data can be formed into I age groups each with span equal to the fixed period interval, but that each of the I age groups can in turn be partitioned into $H \geq 2$ subgroups of identical span. For simplicity we begin by setting $H = 2$ and considering a 6-age and 3-period array with a dichotomous response variable. This is the natural extension of the $3 \times 3 \times 2$ case we presented earlier in detail. The data for the $6 \times 3 \times 2$ case can be arrayed as in Table 7.

Table 7 here

The labels for age groups and cohorts in Table 7 illustrate the basic structure of the data: Individuals in a given cohort must age through two adjacent age categories to pass from one period to the next. There are 6 ordered age groups, $\{1, 1', 2, 2', 3, 3'\}$, and 10 chronologically ordered cohorts, $\{1, 1', 2, 2', 3, 3', 4, 4', 5, 5'\}$. Those cohorts labelled with primes appear only with age groups labelled with primes. Thus, except for the common periods, we can separate Table 7 into two tables each resembling Table 1.

Allowing for primes on the subscripts as in Table 7, the basic age-period-cohort model of expressions (2)-(3) is applicable here:

$$\Omega_{ijk} = W + W_{1(i)} + W_{2(j)} + W_{3(i-j+J)}, \text{ where } i - k = j - J,$$

$$(67)$$

$$\Omega_{i'jk'} = W + W'_{1(i')} + W_{2(j)} + W'_{3(i'-j+J)}, \text{ where } i' - k' = j - J,$$

with

$$\sum_i W_{1(i)} + \sum_{i'} W'_{1(i')} = \sum_j W_{2(j)} = \sum_k W_{3(k)} + \sum_{k'} W'_{3(k')} = 0 \ , \tag{68}$$

and we have again to consider the estimability question.

It might be supposed that with age groups half the length of the period intervals the age-period-cohort identification problem disappears. But not only does this identification problem remain, there is an additional, age-cohort dependency. Inspection of interaction contrasts in 2 × 2 subtables, as described for the 3 × 3 × 2 case, determines the estimable contrasts. For age we can estimate the 3 linear contrasts

$$\begin{aligned} &W_{1(1)} - 2W_{1(2)} + W_{1(3)} \ , \\ &W'_{1(1')} - 2W'_{1(2')} + W'_{1(3')} \ , \\ &(W_{1(1)} - W'_{1(1')}) - (W_{1(3)} - W'_{1(3')}) \ . \end{aligned} \tag{69}$$

For period we can estimate $W_{2(2)}$, as in the 3 × 3 × 2 case, because there are still only 3 periods. For cohort we can estimate the 7 linear contrasts

$$\begin{aligned} &W_{3(k)} - 2W_{3(k+1)} + W_{3(k+2)} && \text{for } k = 1,2,3, \\ &W'_{3(k')} - 2W'_{3(k'+1)} + W'_{3(k'+2)} && \text{for } k' = 1',2',3', \\ &(W_{3(1)} - W'_{3(1')}) - (W_{3(2)} - W'_{3(2')}) \ . \end{aligned} \tag{70}$$

Thus, in the 6 × 3 × 2 case we can estimate only nonlinear effects in age, period and cohort, and an identification specification is necessary. This time we are short two parameters for the age effects and cohort effects, and short one parameter for the period effects. As in the case with equally spaced age and period intervals, the identification problem can be resolved by imposing linear restrictions on the effects. However, not just one, but <u>two</u> restrictions are necessary to identify all of the effects. Moreover, the two restrictions can not be placed arbitrarily. The possibilities are these: (a) Two restrictions can be placed on the age effects; (b) two restrictions can be placed on the cohort effects; (c) one restriction can be placed on each of two separate dimensions (e.g., one on age, the other on period). Placing two restrictions on period effects does not identify all age and cohort effects. In the 6 × 3 × 2 case, one restriction on the period effects suffices to identify all of them, as in the 3 × 3 × 2 case, and hence a second restriction on the period effects is wasted. Indeed, for the 6 × 3 × 2 case, two restrictions on period effects reduces the problem to that of fitting an age-cohort-response array. When the time spanned by two age groups equals the period intervals, however, there is a linear dependency between age and cohort apart from the age-period-cohort dependency; hence the necessity of two identifying restrictions and of placing one of them on age or cohort effects. Similarly, to identify all parameters in an age-cohort model when collapsing over period in the 6 × 3 × 2 case, one identifying restriction must be made.

Estimation of expected cell values or parameters in the $6 \times 3 \times 2$ case introduces no new problems. To use iterative proportional fitting, (a) unfold the frequency table into a four-dimensional incomplete array allowing for 10 cohorts, 6 ages, 3 periods and a dichotomous response variable, and (b) estimate the parameters subject to the identification specification (which will be over-identifying if three or more restrictions are made) following the general procedure illustrated previously for the $3 \times 3 \times 2$ case. To use the Newton-Raphson algorithm first resolve the identification problem by employing two or more linear restrictions (placing at least one of them on age or cohort effects) and then form a design matrix to estimate the 15 (or fewer) free parameters, as described earlier for the $3 \times 3 \times 2$ case.

Degrees of freedom are calculated by subtracting the number of free parameters from the number of log-odds on the response variable. For the age-period-cohort model in the $6 \times 3 \times 2$ case, 18 log-odds are fitted using 15 free parameters and there are 3 degrees of freedom for the just-identified age-period-cohort model. Degrees of freedom for this model and various special cases can be read from Table 6, setting $H = 2 = L$.

For $H > 2$ the identification problem is still two-fold. First, there is an age-period-cohort dependency which requires one identification specification to resolve. Second, there is an age-cohort dependency which requires $H - 1$ identification specifications on age and/or cohort parameters to resolve. Estimation of estimated cell values and parameters in the general $HI \times J \times L$ case encounters no special problems. Table 6 lists degrees of freedom for the full and reduced models.

Irregularly Spaced Periods

When we have data for a categorical response variable broken down by age for several irregularly spaced periods we still might wish to separate the effects of age, period and cohort. While the determination of age and period is no problem here, cohorts become mixed up as we go from one period to another. To consider an example, suppose we have four ten-year age groups and 3 surveys separated by 8 and 12 years respectively, as depicted in Figure 1. The cohorts in period 1 can be linked directly with those in period 3, but for period 2 they get mixed up. Thus, the cells for age group 2 and period 2 correspond to individuals who are in both cohorts 2 and 3. What we must do is allocate these individuals to two cohorts.

Figure 1 here

This problem can be viewed as a generalization of the one studied by Chen and Fienberg (1976) involving totally mixed up frequencies in the analysis of contingency tables. Here it is not that the frequencies themselves are mixed up, but rather that we do not know which value the subscript for one of the parameters in our model takes, in certain cells. Dempster et al. (1977) propose a general approach to problems such as this one which assumes that the "mixing up" is at random, and suggest the repeated application of a two-stage estimation procedure. The first stage involves the allocation of the mixed up data in order to estimate the complete-data sufficient statistics. The second stage involves the solution of a set of equations based on setting the estimated minimal sufficient statistics equal to their expected values. Such a procedure could be devised for the present problem.

Unfortunately, the mixing up of data from different cohorts has not taken place at random in our problem, because the oldest individuals from one cohort are mixed with the youngest ones from the next cohort. As a result, the Dempster et al. approach is not really applicable, although it may serve as a good first approximation. Further work needs to be done before the practitioner can adequately be able to handle data with irregular age groups or irregularly spaced periods.

EXAMPLE

This section illustrates the foregoing discussions of identification, estimation, the nature of the response variable and model fitting, and it illustrates reasoning in aid of model formulation, over-identifying restrictions and interpretation of the results.

The phenomenon we consider is the reported formal educational attainment of cohorts of white males. The data are from the U.S. Census. We argue that education is nondecreasing with age. Were it not for population dynamics, sampling and measurement errors, educational attainment as reported for cohorts in the Decennial Censuses would be constant for most adults and would increase somewhat for the others; it would decrease for nobody. Inspection of Census data indicates that there is some variation over time in the reported educational attainment of cohorts. We provide reasons related to population dynamics, sampling and measurement errors justifying the estimation of age-period-cohort models of educational attainment in order to secure the most accurate possible estimates of cohort educational differentials. Study of the estimated full models suggests that reduced models are appropriate for the data we consider.

A number of points are worth noting at the outset. First, because educational attainment is constant for most individuals after early adulthood, the estimation of cohort contrasts is informative; differences between the educational attainments of cohorts do not fluctuate much over the lives of the individuals in them. Therefore, provided there are good reasons for including age and period, and good reasons for the placement of the over-identifying restrictions, educational attainment is a highly appropriate phenomenon to have selected for illustrative purposes. This is not to say that possible fluctuations over time between cohorts rule out the use of age-period-cohort models or the interpretation of cohort contrasts estimated from reduced models. Rather, we have sought to use an example for which the estimation of cohort contrasts is widely acceptable.

Second, it may be possible to incorporate formally the nondecreasing aspect of educational attainment into a model, with consequent changes in the form of the model, interpretation of its effects, goodness of fit, and degrees of freedom. We have not attempted such a modelling effort, however, since our primary purpose is to illustrate the discussion of preceding sections.

Third, as we shall see below, subsets of the data are about as consistent with an age-period model as with a cohort or an age-period-cohort model. We suspect this may happen frequently in other substantive contexts. Since our understanding of educational change is in terms of cohorts, we have no difficulty in deciding which class of models to reject. This emphasizes, once again, the role of substantive considerations in understanding data and models.

Fourth, our use here of an age-period-cohort specification is akin to estimating measurement models separately from structural models in the analysis of covariance structures (Jöreskog, 1973). Ideally, both kinds of models are estimated simultaneously. Thus, if the goal is to model cohort educational differentials, that task may best be accomplished simultaneously with the estimation of educational attainment. In our example no variables other than age, period and cohort determine educational attainment. We regard our limited effort as defensible, however, for two reasons: What we claim the age and period dimensions represent in our measurement model is probably orthogonal to the substantive factors which account for cohort educational differentials, so there would not be much gain in a joint modelling effort. In addition, except as noted below, age and period effects turn out to be negligible, and knowing this should greatly simplify any effort to model cohort educational differentials substantively.

Data

The data for this example are from the Decennial Censuses of 1940, 1950, 1960 and 1970, and pertain only to white males. Working from published tabulations for each Decennial Census, we have constructed the education distribution of the adult population conditional on age. This gives four age-specific education distributions. Thus, the data are separated by decades, which are the period intervals in the models we apply to these data. We have specified age in five-year groups from 20-24 up to an open-ended 75+ category. This defines cohorts in five-year groups also, but there are only decennial readings on the cohorts. That the final age category is open-ended means there is minor confounding of membership in some cohorts, but various analyses we have carried out in a different context suggest that the confounding has no appreciable effect on the results. Figure 2 gives the concordance of cohorts with ages and years.

<u>Figure 2 here</u>

Formulation of the Dependent Quantity

The form of the analysis, the hypotheses, and the conclusions all depend to some extent on the formulation of the dependent quantity--educational attainment. Three strategies seem reasonable. First educational attainment can be scaled by number of grades completed. The second alternative is to treat education as a polytomous variable. The third is to treat it as a series of continuation ratios. The latter pair of alternatives was discussed earlier in general terms.

Scaling educational attainment by grades of school completed is felicitous for many purposes, but is perhaps too restrictive in the present context. The felicity is that scaling education allows it to be treated within the linear model framework, using covariance-based methods of estimation. In the context of three-way cohort models, however, the scaling of education may be too restrictive for the nature of the effects to be studied, because it requires that all age effects, all period effects and all cohort effects be present in the same way and same amount for each increment in the educational attainment sequence. (This same point also applies to models which treat education as ordinal, but fit only a single age-period-cohort function (Bock, 1975, pp. 541-6)). Such constancy may be present, but it is appropriate to determine this fact first, rather than simply to assume it.

Treating education as a polytomous variable is attractive inasmuch as it requires no scaling assumptions. But this virtue is also a defect: Educational attainment is cumulative. Someone who has completed high school is generally conceded to have a greater quantity of education than someone who has not completed high school, other things being equal. It is desirable to take this ordinality into account, in some fashion. Treating education as a polytomy does not do so, and it also fits only a single set of age, period and cohort effects to the education frequencies.

Describing educational attainment as a series of continuation odds is the most useful of the three alternatives for summarizing education, and the one we use here. We indicate why after describing the odds. We have partitioned the education distribution in each of the censal years into six categories:

E_1: grammar school not completed;

E_2: grammar school completed;

E_3: high school not completed;

E_4: high school completed;

E_5: college not completed;

E_6: college (or more) completed.

These categories are justified on the ground that they separate increments in education in accordance with the completion of levels of schooling recognized as socially important. From these categories the following continuation odds may be formed by summing frequencies from the indicated segments of the education distribution:

$(E_2 + E_3 + E_4 + E_5 + E_6)/E_1$ = the odds of completing grammar school;

$(E_3 + E_4 + E_5 + E_6)/E_2$ = the odds of going to high school for those completing grammar school;

$(E_4 + E_5 + E_6)/E_3$ = the odds of completing high school for those entering;

$(E_5 + E_6)/E_4$ = the odds of entering college for those completing high school;

E_6/E_5 = the odds of completing college for those entering.

These odds are asymptotically independent. Therefore, it is possible to estimate separate three-way logit specifications for each of the odds allowing for differential age, period and cohort effects, for each increment in schooling. This is one advantage of using continuation ratios. A second advantage is that it focusses attention on those odds which are substantively most important. For example, if education were treated as a polytomous variable and its ordinal property ignored, we might form ratios between specific categories, such as E_2/E_1, E_3/E_2, E_4/E_3, E_5/E_4 and E_6/E_5. With the exception of the last ratio, these odds are not identical to, nor are they derivable from, the continuation odds. Moreover, they are uninteresting because they are too restrictive: Knowing the odds of _just_ completing grammar school is for some purposes not as helpful as knowing the odds of completing _at least_ grammar school. Thus, forming continuation odds takes advantage of the ordinal nature of educational attainment in a way that forming odds on education as a nominal variable can not. In this example we model continuation log-odds.

Arrangement of the Data

We have compiled the data from the four decennial age-specific education distributions in five tables. In each of these tables the education frequencies are dichotomized consistent with the continuation odds specified above, and are categorized by age and year. Using these frequencies to compute conditional continuation proportions leads to Tables 8-12. We fit logistic response models which include age, period and cohort effects to the frequencies which underlie Tables 8-12. Below, we specify the form of the three-way model fitted.

Tables 8-12 here

Age-Period-Cohort Specification

We fit logistic response models to five continuation arrays, each of dimension 12 × 4 × 2, using the Newton-Raphson algorithm. Consistent with our discussion in the section on differently spaced age groups and period intervals, we note that each of the arrays breaks into two separate 6 × 4 × 2 subarrays with shared periods. It follows immediately that there are 12 ages, 4 periods, 18 cohorts and two sources of under-identification--an age-period-cohort dependency and an age-cohort dependency. We resolve both identification problems by equating the effects of age groups 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, and we then estimate 6 independent age effects, 3 independent period effects and 17 independent cohort effects. The model is over-identified and has (48 - 6 - 3 - 17 - 1) = 21 degrees of freedom. The justification of this specification is as follows.

Justification

The most desirable modality for summarizing the distribution of educational attainments over time is the birth cohort (Duncan, 1968). Formal education takes place during childhood and adolescence, and for the vast majority of the population is complete by early adulthood. The quantity of education a person gains formally is irreversible. Thus, problems of procedure and error aside, measuring a person's educational attainment at any point as an adult should suffice to determine that person's level of schooling. Most studies of cohort educational attainment have done this; they have measured the level achieved by taking the results of a single sample survey or census, classifying respondents by age, and then imputing educational differences for the age groups to birth cohorts. This procedure is sound if the errors in the data are randomly distributed, if there is no education-mortality-age interaction, if the sample of each cohort is representative, and if the population is closed. We can gain some insight into the soundness of the procedure by pooling data across censuses. (Of course, restricting our attention to censuses only, as we do here, means that our conclusions are largely limited to census-derived estimates of educational attainment.) The only way we can determine whether phenomena associated with the point at which data on educational attainment are collected affect attainment estimates is to obtain data which contain period variation. In addition, repeated cross-sections provide leverage for determining whether certain phenomena, hypothesized to be associated both with education and age, actually do affect cohort differentials in educational attainment. We have justified pooling the data and including cohorts in the specification; we turn next to age and period.

With respect to age we can apply the general argument that recall accuracy decays as a function of time, and that social desirability helps determine what is recalled, so that the older a cohort gets, the higher its education appears. By this argument, the lowest educational category will lose individuals and the highest will gain. The argument implies that for the first continuation odds (grammar school completion) the age coefficients should increase for older age groups, controlling cohort. For the other continuation odds the argument is not decisive, but we would not be surprised to see age increases for high school and college graduation, because of heaping at these important levels of educational attainment.

A second factor with implications for the age coefficients of a model which includes cohort (and period) is that there is an age-education-mortality interaction for white males. For most of adult life, the more schooling you have, the less likely you are to die. However, the education differentials in mortality rates decline with increasing age, and for those ages 75 and over there is a slight reversal. Analysis of data presented by Kitagawa and Hauser (1973, p. 27) suggests that at least for the first four continuation levels (their data do not break out college completion) the age effects should increase through age group 70-74, and should decline for the last (75+) age group.

A third reason for controlling age is that the data to be analyzed include young adults who have not completed their education. Particularly for the continuation ratios associated with higher levels of education, we expect that controlling cohort will lead to increments in the age effects between ages 20-24 and 25-29.

Other reasons can be advanced for controlling age, but the ones we have mentioned seem the most important. Taken together, the age-related hypotheses work in the same direction: For some if not all continuation levels education should increase with age, controlling cohort.

There are several reasons for controlling period. First, Census Bureau procedure has changed over time (Shryock and Siegel, 1973). In 1940 educational attainment was a 100% item. In 1950 it was asked of a

20% sample, by direct enumeration. In 1960 enumerators left a questionnaire with a 25% sample, and those sampled mailed their questionnaires back (in the densely settled areas). In 1970 a mail-out and mail-back questionnaire was used for the 20% sample from which education was ascertained. These differences in the administration of the censuses could have an effect on the results obtained. Enumerators may introduce systematic biases into responses. A mailed questionnaire avoids this source of error, but may introduce others.

A second reason for controlling period is that the published education tabulations for 1940 and 1950 include persons for whom educational attainment was not reported. We have excluded these people from our tables. The 1960 and 1970 published tabulations contain allocated nonrespondents to the education question. This discrepancy in method could result in period differences between 1940 and 1950 on one hand, and 1960 and 1970 on the other. To the extent that nonresponse is distributed over the entire educational distribution, these period effects should be visible in all levels of continuation.

The implications of these two reasons for controlling period are consistent with respect to specific years, but not wholly evident regarding which continuation odds should be affected. If the key procedural difference between censuses for the question we are concerned with is the switch to self-administered questionnaires, and if the consequences of allocation are visible and affect the entire education distribution, then the period effects should be paired (1940, 1950) and (1960, 1970). Presumably self-administration of the education question reduces the social desirability component and therefore reduces the likelihood of upward response bias. If so, the coefficients for 1960 and 1970 should be lower than those for 1940 and 1950, particularly for grammar school completion and perhaps also for high school and college completion. In addition, nonresponse is probably greater at the lower end of the education distribution. If the Census correction for nonresponse is valid, educational attainment in 1960 and 1970 should be lower than in 1940 and 1950--years in which nonresponses were excluded. The consequence of the correction should reduce the 1960 and 1970 effects on grammar school completion; its impact on other continuation odds is less clear.

The reasons for considering age and period effects given here do not make a catalogue. Other possibilities, such as coverage problems and the lack of a closed population, are known to us but have been omitted. In addition, an age-period-cohort model is only one of a number of kinds of models which might be applied to the data which underlie Tables 8-12. The model we use specifies additive age, period and cohort effects on log-odds. It may be that there are plausible reasons not apparent to us for postulating some other kind of model.

We equate the effects of ages 30-34, ... , 55-59 in our age-period-cohort specification. Why? First, since education varies primarily by cohort we want no restrictions on the cohort coefficients. Second, we have a priori grounds for interpreting age effects primarily at the tails of the age distribution. Equating middle age group effects gains identification in a way most consistent with our interests. Third, there is multicollinearity in any fairly unrestricted age-period-cohort model; linear restrictions reduce it. For these last two reasons we have over-identified the model. Since we can reasonably over-identify the age-period-cohort specification using age effects, we have employed no restrictions on the period effects.

Results

To begin with, we have estimated five age-period-cohort models from the continuation data which underlie each of Tables 8-12. The age, period and cohort effects, respectively, are assembled for all continuation levels in Tables 13-15. We present the effects in the form of increments and decrements to the log-odds of schooling continuation.

Tables 13-15 here

Turning to Table 13, we see clear indication of a "completion" effect for the earliest ages, for college graduation. There is a similar but much smaller corresponding effect for college attendance, but there is no evidence to support either the conjecture of increased upward bias in educational reporting with age, or the age-education-mortality interaction described earlier.

Considering the estimated period effects (Table 14) next, we note that those for high school attendance, high school graduation and college attendance (columns 2-4) are perhaps farthest from the hypothesized pattern, while those for grammar school completion and college completion are perhaps closest. In no obvious way are these effects as conjectured. These results provide no basis for supposing that the introduction of mail-back or mail-out-mail-back questionnaires into Census Bureau procedure has had a systematic or marked effect on reported educational attainment. It is, however, noticeable that the 1970 coefficients are with one exception larger than the 1960 coefficients. This could reflect a recent increase in adult education, but in the present context we refrain from pursuing this line of speculation.

Table 15 presents the estimated cohort effects. As noted earlier in this section, the *a priori* rationale for the age-period-cohort specification is that it might provide a way of securing preferred estimates of cohort differentials in educational attainment. An attempt to explain these particular differentials would be premature, however, since the evidence thus far suggests that reduced models are more appropriate than age-period-cohort models for these data. Examination of the goodness-of-fit of reduced models and the over-identified age-period-cohort specification provides additional support for this conclusion.

Table 16 describes the goodness-of-fit of a variety of models, including the over-identified age-period-cohort model we have been discussing. The first row of the table gives the likelihood ratio statistics for each continuation level, for the model we have selected as the baseline. This choice of baseline model is arbitrary but reasonable--we have chosen the simplest model as the basis for comparisons. This model can be described in two ways: In the logit specification it represents fitting the general mean, but not age, period or cohort effects. It also represents a model which fits the margin of the continuation frequencies and the age-period configuration, in the five 12 × 4 × 2 frequency tables which underlie Tables 8-12. Successive rows of Table 16 list the proportionate reduction in the baseline likelihood ratio statistic due to the addition of the specified factors. For example, in the column for grammar school completion, in the row for the age-period-cohort model, we have

$$.999 = [G^2(\text{Baseline}) - G^2(\text{A-P-C})]/[G^2(\text{Baseline})],$$

where G^2 denotes the likelihood ratio statistic. Thus, all likelihood ratio statistics for the fitted models are derivable, and other baselines can be selected, from the information in the table. The degrees of freedom for conditional tests are obtained by appropriate differencing of the degrees of freedom listed in the left-hand margin. For example, the degrees of freedom in the comparison of the age-period-cohort model with the baseline model are 47 - 21 = 26. We indicate degrees of freedom largely for illustrative purposes since, as noted in an earlier section, there is no point in carrying out formal significance tests with these data, and we use the information in Table 16 is useful primarily for describing relative goodness-of-fit.

<u>Table 16 here</u>

The results presented in Table 16 suggest that reduced models fit the data well. The model which allows only for cohort differences accounts for 99% of the lack of fit of the baseline model for grammar school completion and high school attendance, and accounts for 95% of the lack of fit of the baseline model for high school completion. We have already inspected the period and age effects in the age-period-cohort specification for these continuation levels and found these effects neither to conform to the conjectured pattern nor to manifest any marked tendencies whatever. A similar conclusion holds for the cohort-period and cohort-age models listed in Table 16, for the first three continuation ratios. Given the relatively good fit of the cohort-only specification and the inconsistency of the age and period effects with the patterns we conjectured, we consider the cohort-only model to be optimal for these data, regardless of what the conditional likelihood ratio statistics for age and/or period effects would show. This conclusion, however, does not close the door on further attempts to model these data. In particular, the fit of the cohort-only model, while good, can nonetheless be improved. Also, other analysts may be able to explain the age and period effects or posit different models which include cohort and specific, substantively meaningful interactions.

Turning next to the columns in Table 16 pertaining to the final two continuation levels, we see that the relative degree of fit of the cohort-only model deteriorates markedly. For the college attendance continuation ratio, the model accounts for 78% of the lack of fit in the baseline model, and for the college completion continuation ratio, the model accounts for only 49% of the lack of fit of the baseline model. The age-period-cohort model fits the data for these last two continuation ratios extremely well, accounting for 99% of the lack of fit of the baseline model. However, we have already seen that the pattern of period effects does not conform to expectations and is not marked in any event. Moreover, inspection of the age effects for these two continuation ratios showed that the only clear pattern was associated with a completion effect for early adulthood. This suggests that a model somewhere between the cohort-only and age-period-cohort models in complexity may be optimal. In particular, a model which allows for cohort effects and for age effects between ages 20-24, 25-29, and 30+ might fit the data well. As may be seen in Table 16, such a model fits the data very well. For the college attendance continuation ratio, the model accounts for 94% of the lack of fit in the baseline model, and for college graduation, this cohort and highly restricted age model accounts for 98% of the lack of fit of the baseline model. For the last continuation ratio, it is clear that allowing only for a completion effect due to age and for cohort differences fits the data just as well as the age-period-cohort model. For the college attendance continuation ratio, however, this model fits less well than the age-period-cohort model. If we were relying exclusively on measures of fit, this result might suggest the need for including in the model more than just cohort plus two independent age effects. However, since we have already seen that the period and age effects (aside from the modest completion effect with age for this continuation ratio) are small and run counter to our conjectures, we prefer to stop the fitting procedure at this point. Moreover, even if we allow age to enter into the model in more detail and improve the fit slightly, we find a pattern to the age effects which we regard as uninterpretable. Thus, we conclude that for the last two continuation ratios the appropriate specification is one which allows for cohort differences and for a completion effect in early adulthood.

From Table 16 it can be seen that an age-period model (i.e., excluding cohort) in some instances fits the data quite well. This does not necessarily mean that this model is ever preferable to those we have settled on. Since it is through cohort replacement that education distributions change, it is essential to include the cohort classification in any analysis of these data. The meaning of age and period effects can only be in terms such as those we have used, that is, as minor adjustments to the general pattern of cohort differentials. Thus, we reject the age-period model on conceptual grounds.

We consider next the results of estimating the cohort-only model for the first three continuation ratios, and the age-cohort model (with equality restrictions on the coefficients of ages 30+) for the last two continuation ratios. Tables 17-18 present the age and cohort effects estimated for these models. As before, the cohort effects for all continuation ratios are collected in one table, and the age effects for the last two continuation ratios are presented in a single table.

<u>Tables 17 and 18 here</u>

As Table 17 shows, the age effects for the two highest continuation ratios are as hypothesized. The likelihood of continuing to a given educational level increases monotonically to age 30, and is thereafter constrained to be constant. As expected, and as seen originally in Table 13, the aging increment in early adulthood is greatest for college completion.

The pattern of cohort effects presented in Table 18 is complex, and the explanation of it can not be the subject here of extended discussion. Major research on continuation ratios such as these is currently under way (Mare, 1976). It will perhaps suffice for our purposes to note first that the patterns of the cohort contrasts are similar to those discussed by Duncan (1968). Since Duncan's presentation mode differs from ours, small discrepancies are to be expected. In addition, we would expect differences because we have obtained estimates of cohort attainment using pooled cross-sections, whereas Duncan relies on single measures for each cohort, and uses Current Population Survey data as well as census data (Duncan, 1968, p. 655).

Having commented on the parameters of the parsimonious models we consider most appropriate for the data, we assess overall goodness-of-fit. As discussed in an earlier section, because of the asymptotic independence of the different levels of continuation odds we can add likelihood ratio statistics across levels of continuation to obtain an indication of overall fit. Doing so we find that the fit of the set of parsimonious models we have selected (listed in Tables 17-18) is good. The proportionate reduction in the lack of fit of the overall baseline model for all continuation ratios (which allows only for differences in rates across the continuation levels) by the set of parsimonious models is .985. This compares well with .998 for the proportionate reduction of the lack of fit of the overall baseline model by the overall age-period-cohort model (which does not constrain coefficients across levels of continuation). If we estimate an age-period-cohort model which restricts the age, period, and cohort effects to be equal for all levels of continuation but allows for differences in the continuation rates between levels, we find that the proportionate reduction in the lack of fit of the overall baseline model is .595. This result reinforces one of the initial premises of this analysis--that it is more informative to allow the effects of age, period and cohort to vary across levels of education than to restrict them a priori by the form of the dependent quantity, a point which inspection of the estimated coefficients has also made obvious.

Discussion

We have tried to illustrate the nature of the reasoning we think appropriate for use with age-period-cohort models, to follow through with a consistent data analysis, and to illustrate a number of results presented in earlier sections. There is nothing special about reasoning with age-period-cohort models; all it requires is plausible and presumably non-trivial ratiocination about the inclusion of age, period and cohorts; prior expectations about the nature of at least some of their effects; and an identification specification. The prior reasoning may be susceptible to varying degrees of formalization, but the requirement that it exist is neither more nor less than for any other kind of modelling effort. The requirement of an identification specification is not unique to age-period-cohort models, and in other contexts linear restrictions are often made routinely (e.g., treating region of residence as South vs. other, marital status as currently married vs. other, race as white vs. other). Nonetheless, the identification specification does require justification.

In our example we suggested that pooling data over cross-sections might lead to improved estimates of cohort differentials in educational attainment and explained why, and then went on to justify the age-period-cohort specification, conjecturing certain age and period effects. We did not specify a pattern of cohort contrasts in advance since our concern was not to estimate a model explaining them, but rather to secure improved estimates of the differentials themselves. We made an over-identifying specification so that the model would have unique implications for fitting the data, and we defended our equality restrictions. We estimated the age-period-cohort specification, concluded that the data did not support most of our conjectures about age and period effects, and settled on reduced models.

It would be appropriate to explain the cohort contrasts estimated by the reduced models we selected. In so doing it might be desirable to incorporate the explanatory factors into the parsimonious models we stopped with, but a decisive judgment can not be reached separately from considerations about the form of the explanatory model and data availability.

We indicated the usefulness of forming continuation ratios for the study of educational attainment and partially demonstrated their value by showing that different levels of educational continuation require different models. Our example used 5-year age groups and 10-year period intervals, and we had to make an identification specification which took account of the age-cohort dependency as well as the age-period-cohort dependency. We summarized the goodness-of-fit of our models by computing a measure of overall relative goodness-of-fit, taking advantage of the separability of the models of the continuation odds.

FOOTNOTES

[1]Goodman (1975) discusses the estimation of age, period, and cohort models for discrete data drawing largely on Goodman (1972), but does not cover as many topics as we do, and does not provide as much detail as we do on the points we all make. Our work on this chapter began in 1974, with the recognition that most of the major points were already known (Fienberg, 1972; Mason et al., 1973), and we reported on it in abbreviated form in 1976 (Fienberg and Mason, 1976).

[2]Each of the sampling models described below is of exponential family form, and the logarithms of the response probabilities or odds are technically referred to as the natural parameters for exponential-family distributions (Andersen, 1974; Dempster, 1971). Since we are interested in the effects of age, period, and cohort on the response probabilities it then turns out to be convenient to deal with the difference $\log(P_{ijk|1}) - \log(P_{ijk|2})$, which is also known as the logit (Goodman, 1975b).

[3]The basic structure for the probabilities in the $I \times J$ array is unbalanced with respect to cohort, in that there is only one probability corresponding to cohorts 1 and $I + J - 1$, two probabilities corresponding to cohorts 2 and $I + J - 2$, etc. As a result of this imbalance, and the use of symmetric constraints in expression (3), the parameter W is not the mean of the $\{\Omega_{ijk}\}$. Since the constraints in (3) are to a large extent arbitrary, one might well choose to replace $\Sigma_k W_{3(k)} = 0$ by the weighted constraint $\Sigma_k \omega_k W_{3(k)} = 0$, where the weight, ω_k, is proportional to the number of probabilities corresponding to the k-th cohort. The

use of the weighted constraint would allow us to interpret W as a grand mean, whereas the use of (3) implies only that W is a normalization constant. Since we are interested in the effects of age, period, and cohort on the response variable, W is merely a nuisance parameter and we use the symmetric constraints in (3) for simplicity.

[4]The idea of marginality (Nelder, 1974) is crucial to a technical understanding of estimability of parameters and degrees of freedom for model fitting. It is related to the geometric representation of linear models, as opposed to the algebraic representation most often adopted in social science applications.

[5]As Goodman (1975a) notes, fitting [AP], [AC] or [PC] is equivalent to fitting [APC], for the age-period-cohort model.

[6]It is possible to estimate the expected frequencies directly for reduced models with only one set of effect parameters (Bishop, Fienberg and Holland, 1975, pp. 74-79). There is no practical advantage in exploiting this fact since the iterative proportional fitting algorithm applied to such reduced models will converge in one cycle.

[7]Pullum (1978, pages forthcoming) asserts that "it is difficult to see how any researcher could a priori argue for setting a particular pair of effects to exactly the same value." He proposes setting an effect equal to a nonzero constant, and indicates a method of empirically determining the value of the constant. Setting an effect equal to a constant is in general no more defensible than equating two effects. Moreover, the procedure Pullum outlines for determining the constant yields an acceptable value only insofar as the procedure can be justified

on the basis of a priori reasoning and knowledge in the specific context in which it is applied.

[8]Continuation fractions are commonly presented as probabilities, $\sum_{\ell' > \ell} m_{ijk\ell'} \,/\, \sum_{\ell} m_{ijk\ell}$, in which form they are usually known as continuation ratios. We discuss (the negative logarithms of) continuation odds instead of continuation ratios for the reasons given in our specification of model (2)-(3).

REFERENCES

Altham, P. M. E.

1976 "Discrete Variable Analysis for Individuals Grouped into Families." Biometrika 63:263-70.

Andersen, A. H.

1974 "Multidimensional Contingency Tables." Scandanavian Journal of Statistics 1:115-27.

Baltes, P. B., Cornelius, S. W., and Nesselroade, J. R.

1976 "Cohort Effects in Developmental Psychology: Theoretical and Methodological Perspectives." Pp. - in W. A. Collins (ed.), Minnesota Symposium on Child Psychology, Volume II. Minneapolis: Institute of Child Development, University of Minnesota.

Bartlett, M. S.

1935 "Contingency Table Interactions." Journal of the Royal Statistical Society, Supplement 2:248-52.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W.

1975 Discrete Multivariate Analysis: Theory and Practice. Cambridge, Massachusetts: The MIT Press.

Blalock, H. M.

1966 "The Identification Problem and Theory Building: The Case of Status Inconsistency." American Sociological Review 31:52-61.

Bock, R. D.

1975 Multivariate Statistical Methods in Behavioral Research. New York: McGraw-Hill.

Bock, R. D., and Yates, G.

1973 "MULTIQUAL, Loglinear Analysis of Nominal or Ordinal Qualitative Data by the Method of Maximum Likelihood: a FORTRAN IV Program." Chicago, Illinois: National Educational Resources.

Chen, T., and Fienberg, S. E.

1976 "The Analysis of Contingency Tables with Incompletely Classified Data." Biometrics 32:133-44.

Cohen, J. E.

1976 "The Distribution of the Chi-Squared Statistic Under Clustered Sampling from Contingency Tables." Journal of the American Statistical Association 71:665-70.

Darroch, J. N., and Ratcliff, D.

1972 "Generalized Iterative Scaling for Loglinear Models." Annals of Mathematical Statistics 43:1470-80.

Dempster, A. P.

1971 "An Overview of Multivariate Analysis." Journal of Multivariate Analysis 1:316-46.

Dempster, A. P., Laird, N. M., and Rubin, D. B.

1977 "Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion)." Journal of the Royal Statistical Society, Series B 39: in press.

Duncan, B.

1968 "Trends in Output and Distribution of Schooling." Pp. 601-72 in E. B. Sheldon and W. E. Moore (eds.), Indicators of Social Change. New York: Russell Sage Foundation.

Fay, R., and Goodman, L. A.

1973 "Everyman's Contingency Table Analyzer." Department of Sociology, University of Chicago.

Fienberg, S. E.

1972 "The Analysis of Incomplete Multi-way Contingency Tables." Biometrics 28:177-202.

Fienberg, S. E., and Mason, W. M.

1976 "Identification and Estimation of Age, Period, and Cohort Effects in the Analysis of Discrete Archival Data." Proceedings of the Social Statistics Section, American Statistical Association, Part I, pp. 133-7.

Fienberg, S. E.

1977 The Analysis of Cross-Classified Data. Cambridge, Massachusetts: The MIT Press.

Glenn, N. D.

1976 "Cohort Analysts' Futile Quest: Statistical Attempts to Separate Age, Period and Cohort Effects." American Sociological Review 41:900-4.

Goodman, L. A.

1972 "Some Multiplicative Models for the Analysis of Cross Classified Data." Pp. 649-96 in L. Le Cam, J. Neyman, E. Scott (eds.), Proceedings of The Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume I. Berkeley, California: University of California Press.

Goodman, L. A.

1975a "A Note on Cohort Analysis Using Multiplicative Models and the Modified Multiple Regression Approach." Manuscript, Department of Sociology, University of Chicago.

Goodman, L. A.

1975b "The Relationship Between Modified and Usual Multiple Regression Approaches to the Analysis of Dichotomous Variables." Pp. 83-110 in D. R. Heise (ed.), Sociological Methodology 1976. San Francisco: Jossey-Bass.

Graybill, F. A.

1976 Theory and Application of the Linear Model. North Scituate, Massachusetts: Duxbury Press.

Greenberg, B. G., Wright, J. J., and Sheps, C. G.

1950 "A Technique for Analyzing some Factors Affecting the Incidence of Syphilis." Journal of the American Statistical Association 251:373-99.

Grizzle, J. E., Starmer, C. F., and Koch, G. G.

1969 "Analysis of Categorical Data by Linear Models." Biometrics 25:489-504.

Haberman, S. J.

1973 "C-TAB, Analysis of Multidimensional Contingency Tables by Loglinear Models: a FORTRAN IV Program." Chicago, Illinois: National Educational Resources.

Haberman, S. J.

1974 The Analysis of Frequency Data. Chicago: University of Chicago Press.

Haberman, S. J.

1976 "Book Review of Discrete Multivariate Analysis: Theory and Practice." Annals of Statistics 4:817-20.

Haberman, S. J.

1978 Untitled manuscript on the analysis of frequency data. Department of Statistics, University of Chicago. Forthcoming.

Hall, R. E.

1971 "The Measurement of Quality Change from Vintage Price Data." Pp. 240-71 in Z. Griliches (ed.), Price Indexes and Quality Changes. Cambridge, Massachusetts: Harvard University Press.

Hope, K.

1971 "Social Mobility and Fertility." American Sociological Review 36:1019-32.

Hope, K.

1975 "Models of Status Inconsistency and Social Mobility Effects." American Sociological Review 40:322-43.

Jöreskog, K. G.

1973 "A General Method for Estimating a Linear Structural Equation System." Pp. 85-112 in A. S. Goldberger and O. D. Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Kish, L., and Frankel, M. R.

1974 "Inference from Complex Samples (with discussion)." Journal of the Royal Statistical Society, Series B 36:1-37.

Kitagawa, E. M., and Hauser, P. M.

1973 Differential Mortality in the United States. Cambridge, Massachusetts: Harvard University Press.

Knoke, D., and Hout, M.

1974 "Social and Demographic Factors in American Political Party Affiliation, 1952-72." American Sociological Review 39:700-13.

Knoke, D., and Hout, M.

1976 "Reply to Glenn." American Sociological Review 41:905-8.

Koch, G. G., Freeman, D. H., and Freeman, J. L.

1973 "Some Useful Strategies in the Multivariate Analysis of Data from Complex Surveys." Proceedings of the Social Statistics Section, American Statistical Association, pp. 8-17.

Mare, R. D.

1976 "Growth and Distribution of Schooling in the Twentieth Century United States." Ph.D. Dissertation Proposal. Department of Sociology, University of Michigan.

Mason, K. O., Mason, W. M., Winsborough, H. H., and Poole, W. K.

1973 "Some Methodological Issues in Cohort Analysis of Archival Data." American Sociological Review 38:242-58.

Mason, W. M., Mason, K. O., and Winsborough, H. H.

1976 "Reply to Glenn." American Sociological Review 41:904-5.

National Assessment of Education Progress

1972 Reading and Literature: General Information Yearbook. Report 02-GIY. Washington, D.C.: U.S. Government Printing Office.

Nelder, J. A.

1974 Letter to the Editor on "Analysis of Variance Programs, Least Squares, and Square Tables." Applied Statistics 23:232.

Penick, B. K. E., and Owens, M. W. B. (eds.)

1976 Surveying Crime. Washington, D.C.: National Academy of Sciences.

Pullum, T. W.

1975 Measuring Occupational Inheritance. New York: Elsevier.

Pullum, T. W.

1978 "Parametrizing Age, Period, and Cohort Effects, with Application to U.S. Delinquency Rates, 1964-1973." Pp. - in K. F. Schuessler (ed.), Sociological Methodology 1978. San Francisco: Jossey-Bass.

Schaie, K. W.

1977 "Quasi-Experimental Research Designs in the Psychology of Aging." Pp. 39-57 in J. W. Birren and K. W. Schaie (eds.), Handbook of the Psychology of Aging. New York: Van Nostrand-Reinhold.

Shryock, H. S., Siegel, J., and Associates

1973 The Methods and Materials of Demography. Second Printing (rev.). U.S. Bureau of the Census. Washington, D.C.: U.S. Government Printing Office.

Shuster, J. J., and Downing, D. J.

1976 "Two-way Contingency Tables for Complex Sampling Schemes." Biometrika 63:271-6.

Winsborough, H. H.

1976 "Discussion." Proceedings of the Social Statistics Section, American Statistical Association, Part I, pp. 138-9.

Yates, F.

1934 "The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes." Journal of the American Statistical Association 29:51-66.

		Period		
		1940	1948	1960
Age Group	20-29	3	1&2	1
	30-39	4	2&3	2
	40-49	5	3&4	3
	50-59	6	4&5	4

Figure 1. An Illustration of Cohort Structure with Irregularly-Spaced Periods

Age		Year 1: 1940	2: 1950	3: 1960	4: 1970
		Cohort			
1:	20–24	7	5	3	1
2:	25–29	8	6	4	2
3:	30–34	9	7	5	3
4:	35–39	10	8	6	4
5:	40–44	11	9	7	5
6:	45–49	12	10	8	6
7:	50–54	13	11	9	7
8:	55–59	14	12	10	8
9:	60–64	15	13	11	9
10:	65–69	16	14	12	10
11:	70–74	17	15	13	11
12:	75+	18	16	14	12

Figure 2. Layout for Estimating Schooling Continuation Logits as Functions of Age, Period and Cohort

TABLE 1

Age by Period Display

		Positive Response Period					Negative Response Period		
		1	2	3			1	2	3
	1	x_{1131}	x_{1221}	x_{1311}		1	x_{1132}	x_{1222}	x_{1312}
Age	2	x_{2141}	x_{2231}	x_{2321}	Age	2	x_{2142}	x_{2232}	x_{2322}
	3	x_{3151}	x_{3241}	x_{3331}		3	x_{3152}	x_{3242}	x_{3332}

TABLE 2

Age by Cohort Display (Positive Response)

		Cohort				
		1	2	3	4	5
	1	x_{1311}	x_{1221}	x_{1131}	–	–
Age	2	–	x_{2321}	x_{2231}	x_{2141}	–
	3	–	–	x_{3331}	x_{3241}	x_{3151}

TABLE 3

Period by Cohort Display (Positive Response)

		Cohort 1	2	3	4	5
Period	1	-	-	x_{1131}	x_{2141}	x_{3151}
	2	-	x_{1221}	x_{2231}	x_{3241}	-
	3	x_{1311}	x_{2321}	x_{3331}	-	-

TABLE 4

Unfolded Four-Dimensional Data Array

Cohort			Positive Response Period 1	Positive Response Period 2	Positive Response Period 3	Negative Response Period 1	Negative Response Period 2	Negative Response Period 3
		1	–	–	x_{1311}	–	–	x_{1312}
1	Age	2	–	–	–	–	–	–
		3	–	–	–	–	–	–
		1	–	x_{1221}	–	–	x_{1222}	–
2	Age	2	–	–	x_{2321}	–	–	x_{2322}
		3	–	–	–	–	–	–
		1	x_{1131}	–	–	x_{1132}	–	–
3	Age	2	–	x_{2231}	–	–	x_{2232}	–
		3	–	–	x_{3331}	–	–	x_{3332}
		1	–	–	–	–	–	–
4	Age	2	x_{2141}	–	–	x_{2142}	–	–
		3	–	x_{3241}	–	–	x_{3242}	–
		1	–	–	–	–	–	–
5	Age	2	–	–	–	–	–	–
		3	x_{3151}	–	–	x_{3152}	–	–

TABLE 5

Constraints Associated with the 3 × 3 × 2 Array

Positive Response

		Period 1	Period 2	Period 3
Age	1	Δ	$-\Delta$	0
	2	$-\Delta$	0	Δ
	3	0	Δ	$-\Delta$

Negative Response

		Period 1	Period 2	Period 3
Age	1	$-\Delta$	Δ	0
	2	Δ	0	$-\Delta$
	3	0	$-\Delta$	Δ

TABLE 6

Information Associated with Age-Period-Cohort Model and Various Reduced Models

Subscripted Logistic Parameters in Model (2)-(3)	Degrees of Freedom[a]	Minimal Sufficient Statistics[b]
None	$(HIJ - 1)(L - 1)$	$\{x_{+++\ell}\}$
Age	$HI(J - 1)(L - 1)$	$\{x_{i++\ell}\}$
Period	$J(HI - 1)(L - 1)$	$\{x_{+j+\ell}\}$
Cohort	$H(I - 1)(J - 1)(L - 1)$	$\{x_{++k\ell}\}$
Age, Period	$(HI - 1)(J - 1)(L - 1)$	$\{x_{i++\ell}\}$, $\{x_{+j+\ell}\}$
Age, Cohort	$H(I - 1)(J - 2)(L - 1)$	$\{x_{i++\ell}\}$, $\{x_{++k\ell}\}$
Period, Cohort	$(HI - H - 1)(J - 1)(L - 1)$	$\{x_{+j+\ell}\}$, $\{x_{++k\ell}\}$
Age, Period, Cohort	$(HI - H - 1)(J - 2)(L - 1)$	$\{x_{i++\ell}\}$, $\{x_{+j+\ell}\}$, $\{x_{++k\ell}\}$

[a]The formulae assume the data can be formed into an $I \times J \times L$ array, with evenly and equally spaced age groups and period intervals. For this case set $H = 1$. If the I age groups have been broken into a fixed number of subgroups for each period interval, set H equal to this number. See text for further discussion.

[b]Each model always includes the totals, $\{x_{ijk+}\}$, implied by the logistic structure, as well as the statistics included.

TABLE 7

Age by Period Display with Two Age Groups
for Each Period Interval

Positive Response

		Period		
		1	2	3
Age	1	x_{1131}	x_{1221}	x_{1311}
	1′	$x_{1'13'1}$	$x_{1'22'1}$	$x_{1'31'1}$
	2	x_{2141}	x_{2231}	x_{2321}
	2′	$x_{2'14'1}$	$x_{2'23'1}$	$x_{2'32'1}$
	3	x_{3151}	x_{3241}	x_{3331}
	3′	$x_{3'15'1}$	$x_{3'24'1}$	$x_{3'33'1}$

Negative Response

		Period		
		1	2	3
Age	1	x_{1132}	x_{1222}	x_{1312}
	1′	$x_{1'13'2}$	$x_{1'22'2}$	$x_{1'31'2}$
	2	x_{2142}	x_{2232}	x_{2322}
	2′	$x_{2'14'2}$	$x_{2'23'2}$	$x_{2'32'2}$
	3	x_{3152}	x_{3242}	x_{3332}
	3′	$x_{3'15'2}$	$x_{3'24'2}$	$x_{3'33'2}$

TABLE 8

Proportion Completing Grammar School, by Age and Year, for White Males in the United States, 1940-1970

	Year			
Age	1940	1950	1960	1970
20-24	.861	.881	.929	.963
25-29	.837	.879	.910	.954
30-34	.806	.863	.892	.936
35-39	.758	.823	.884	.916
40-44	.722	.789	.863	.903
45-49	.667	.742	.820	.894
50-54	.626	.704	.779	.875
55-59	.605	.647	.720	.831
60-64	.588	.596	.672	.786
65-69	.558	.562	.601	.720
70-74	.544	.539	.552	.676
75+	.508	.516	.525	.599

SOURCE: Sixteenth Census of the United States: 1940, Population, Volume IV, Characteristics by Age, Part 1, United States Summary, Table 18; U.S. Census of Population: 1950, Vol. II, Characteristics of the Population, Part 1, United States Summary, Tables 114, 115; U.S. Census of Population: 1960, Vol. 1, Characteristics of the Population, Part 1, United States Summary, Tables 168, 172, 173; Census of Population: 1970, Vol. I, Characteristics of the Population, Part 1, United States Summary--Section 2, Table 199.

TABLE 9

Proportion Attending High School, Conditional on Grammar School Completion, by Age and Year, for White Males in the United States, 1940-1970

	Year			
Age	1940	1950	1960	1970
20-24	.806	.873	.927	.970
25-29	.740	.864	.906	.956
30-34	.680	.825	.880	.938
35-39	.596	.767	.870	.917
40-44	.549	.713	.830	.893
45-49	.506	.640	.766	.881
50-54	.465	.599	.712	.842
55-59	.436	.561	.633	.781
60-64	.412	.524	.586	.726
65-69	.378	.495	.544	.647
70-74	.345	.472	.507	.612
75+	.314	.437	.473	.568

SOURCE: See Table 8.

TABLE 10

Proportion Completing High School, Conditional on High School Attendance, by Age and Year, for White Males in the United States, 1940-1970

	Year			
Age	1940	1950	1960	1970
20-24	.652	.694	.756	.864
25-29	.629	.710	.760	.837
30-34	.603	.692	.716	.813
35-39	.590	.660	.725	.798
40-44	.594	.635	.700	.744
45-49	.608	.627	.659	.740
50-54	.621	.634	.625	.711
55-59	.629	.641	.611	.675
60-64	.652	.644	.605	.646
65-69	.660	.644	.610	.633
70-74	.671	.662	.621	.628
75+	.685	.680	.645	.645

SOURCE: See Table 8.

TABLE 11

Proportion Attending College, Conditional on High School Completion, by Age and Year, for White Males in the United States, 1940-1970

	Year			
Age	1940	1950	1960	1970
20-24	.316	.419	.432	.533
25-29	.381	.411	.465	.499
30-34	.452	.389	.480	.471
35-39	.490	.418	.442	.481
40-44	.487	.474	.412	.486
45-49	.493	.497	.436	.454
50-54	.500	.489	.492	.417
55-59	.490	.489	.528	.434
60-64	.504	.495	.530	.481
65-69	.519	.485	.541	.510
70-74	.518	.499	.551	.501
75+	.518	.510	.542	.508

SOURCE: See Table 8.

TABLE 12

Proportion Completing College, Conditional on College Attendance, by Age and Year, for White Males in the United States, 1940-1970

	Year			
Age	1940	1950	1960	1970
20-24	.270	.215	.250	.244
25-29	.507	.468	.536	.548
30-34	.536	.504	.585	.586
35-39	.544	.520	.568	.616
40-44	.518	.536	.520	.603
45-49	.521	.541	.511	.570
50-54	.532	.520	.514	.524
55-59	.533	.522	.510	.522
60-64	.538	.538	.477	.527
65-69	.536	.535	.468	.532
70-74	.536	.542	.468	.506
75+	.532	.541	.475	.508

SOURCE: See Table 8.

TABLE 13

Age Effects for Schooling Continuation Logits, for White Males, Based on Age-Period-Cohort Loglinear Model[a]

Age Group	Continuation Levels 0-7 vs. 8+	8 vs. 9+	9-11 vs. 12+	12 vs. 13+	13-15 vs. 16+
20-24	-.032	-.023	-.093	-.125	-1.167
25-29	.069	-.003	-.021	.007	-.024
30-59	.107	.014	.056	.062	.236
60-64	.043	.004	.030	.024	.247
65-69	-.018	-.005	.014	.029	.233
70-74	-.060	.016	-.015	.001	.219
75+	-.110	-.003	.029	.003	+.230

[a]The model constrains the coefficients of age groups 30-34,...,50-59 to be identical.

TABLE 14

Period Effects for Schooling Continuation Logits for White Males, Based on Age-Period-Cohort Loglinear Model[a]

	Continuation Levels				
Year	0-7 vs. 8+	8 vs. 9+	9-11 vs. 12+	12 vs. 13+	13-15 vs. 16+
1940	.075	−.147	−.086	−.106	.097
1950	−.023	.029	.012	−.033	−.051
1960	−.072	.008	−.022	.071	−.051
1970	.020	.109	.096	.069	.004

[a]The model constrains the coefficients of age groups 30-34,...,55-59 to be identical.

TABLE 15

Cohort Effects for Schooling Continuation Logits for White Males, Based on Age-Period-Cohort Loglinear Model[a]

Cohort	Continuation Levels				
	0-7 vs. 8+	8 vs. 9+	9-11 vs. 12+	12 vs. 13+	13-15 vs. 16+
Born:					
1946-50	1.995	2.401	1.007	.263	.103
1941-45	1.647	1.977	.727	-.003	.283
1936-40	1.329	1.576	.442	-.158	.165
1931-35	1.002	1.271	.373	-.135	.282
1926-30	.792	.970	.064	-.115	.167
1921-25	.693	.863	.076	-.262	.083
1916-20	.503	.559	-.063	-.415	-.059
1911-15	.186	.169	-.228	-.307	-.062
1906-10	-.052	-.122	-.364	-.077	-.073
1901-05	-.342	-.476	-.413	.050	-.057
1896-1900	-.519	-.657	-.399	.047	-.155
1891-95	-.781	-.822	-.359	.065	-.154
1866-90	-.940	-.979	-.302	.121	-.110
1881-85	-1.018	-1.094	-.265	.088	-.114
1876-80	-1.050	-1.189	-.156	.145	-.042
1871-75	-1.101	-1.310	-.113	.187	-.028
1886-70	-1.126	-1.507	-.021	.255	-.100
-1865	-1.220	-1.630	-.004	.252	-.128

[a]The model constrains the coefficients of age groups 30-34,...,55-59 to be identical.

TABLE 16

Baseline Likelihood Ratio Statistics, and Proportionate Change Thereto, for Various Logit Specifications on Schooling Continuation, for all Levels of Continuation[a]

	Continuation Level				
Model	0-7 vs. 8+	8 vs. 9+	9-11 vs. 12+	12 vs. 13+	13-15 vs. 16+
Baseline (df = 47)[b]	17,220,464.0	20,598,402.0	3,215,005.0	724,218.0	2,039,637.0
Period (df = 44)	.207	.332	.445	.185	.012
Age (df = 36)	.666	.534	.286	.122	.943
Cohort (df = 30)	.988	.991	.948	.783	.491
Period + Age (df = 33)	.972	.978	.822	.294	.967
Period + Cohort (df = 27)	.995	.999	.986	.931	.618
Age + Cohort (df = 20)[c]	.996	.998	.990	.972	.978
Age + Cohort (df = 24)[d]	.996	.996	.985	.947	.977
Age + Cohort (df = 28)[e]	---	---	---	.939	.976
Age + Period + Cohort (df = 21)[d]	.999	1.000	.997	.986	.985

[a]Proportionate improvement is defined as follows. Let $G^2(1)$ be the likelihood ratio statistic for the baseline model. Then the proportionate reduction of $G^2(1)$ due to a hierarchically more complex structure (any model other than the baseline) is computed as $p = [G^2(1) - G^2(2)]/[G^2(1)]$, where $G^2(2)$ is the likelihood ratio statistic for the model with the more complex structure.

[b]Fits general mean to logits.

[c]Coefficients of age groups 30-34, 35-39 equated.

[d]Coefficients of age groups 30-34,...,55-59 equated.

[e]Coefficients of age groups 30-34,...,75+ equated.

TABLE 17

Age Effects for Schooling Continuation Logits, for White Males for College Attendance and College Completion in a Parsimonious Loglinear Model[a]

	Continuation Level	
Age	12 vs. 13+	13-15 vs. 16+
20-24	-.134	-.858
25-29	-.001	.288
30+	.135	.570

[a]Based on a model of the logits which includes cohorts and the specified age effects, and excludes period effects.

TABLE 18

Cohort Effects for Schooling Continuation Logits, for White Males, Over all Continuation Levels, Derived from Parsimonious Loglinear Models[a]

	Continuation Levels				
Cohort	0-7 vs. 8+	8 vs. 9+	9-11 vs. 12+	12 vs. 13+	13-15 vs. 16+
1946-50	1.941	2.493	.997	.420	.108
1941-45	1.695	2.089	.789	.152	.286
1936-40	1.297	1.636	.447	-.045	.151
1931-35	1.022	1.341	.414	-.022	.260
1926-30	.787	1.027	.088	-.046	.141
1921-25	.721	.925	.122	-.194	.053
1916-20	.533	.564	-.059	-.386	-.069
1911-15	.241	.178	-.207	-.281	-.064
1906-10	-.000	-.113	-.331	-.082	-.076
1901-05	-.300	-.473	-.386	.042	-.061
1896-1900	-.495	-.659	-.385	.021	-.162
1891-95	-.786	-.825	-.337	.039	-.160
1886-90	-.931	-1.015	-.321	.059	-.115
1881-85	-1.056	-1.134	-.279	.034	-.126
1876-80	-1.052	-1.254	-.203	.027	-.030
1871-75	-1.175	-1.375	-.143	.077	-.033
1886-70	-1.151	-1.632	-.134	.093	-.043
-1865	-1.294	-1.774	-.073	.091	-.061

[a]Cohort effects for grammar school completion, high school attendance and high school completion are based on a model which fits only cohort effects to the continuation logits. The cohort effects for college attendance and college completion are based on a model which fits effects for age groups 20-24, 25-29, 30-75+, as well as cohort effects.